AD-763 477

THE ANALYSIS OF CONTINGENCY TABLES: A METHODOLOGICAL EXPOSITION

S. Kullback

George Washington University

Prepared for:

Office of Naval Research

21 May 1973

AD763 477

THE ANALYSIS OF CONTINGENCY TABLES - A METHODOLOGICAL EXPOSITION

by

S. Kullback

TECHNICAL REPORT

Serial TR-1116
21 May 1973

THE GEORGE WASHINGTON UNIVERSITY
Graduate School of Arts and Sciences
Econometric Research on Navy Manpower Problems

This report was prepared under the All Volunteer Force R&D Program of the Office of Naval Research under Contract Number N00014-67-A-0214-0016

98

NONE
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) THE GEORGE WASHINGTON UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES ECONOMETRIC RESEARCH ON NAVY MANPOWER PROBLEMS	2a. REPORT SECURITY CLASSIFICATION NONE 2b. GROUP

3. REPORT TITLE

THE ANALYSIS OF CONTINGENCY TABLES - A METHODOLOGICAL EXPOSITION

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

SCIENTIFIC

5. AUTHOR(S) (First name, middle initial, last name)

S. KULLBACK

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
21 May 1973	96	267

8a. CONTRACT OR GRANT NO. N00014-67-A-0214-0016	9a. ORIGINATOR'S REPORT NUMBER(S) TR-1116
b. PROJECT NO. NR 347 024	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research

13. ABSTRACT

The analysis of the information in contingency tables is an aspect of multivariate (multiple variates) analysis with particular application to qualitative or categorical as well as quantitative variables.

The analysis is concerned with counts in multiway cross-classifications or multiple contingency tables. Multiway contingency tables, or cross-classifications of vectors of discrete random variables, provide a useful approach to the analysis of multivariate discrete data.

The method of analysis presented will bring out the various interrelationships among the classificatory variables in a multiway cross-classification or contingency table in many dimensions. The illustration of the procedure is an application to Marine cohort data considering the relation of boot camp completion on home of record, level of education, and race.

The procedure is based on the Principle of Minimum Discrimination Information Estimation, associated statistics and Analyses of Information. General computer programs are available to provide the necessary results for inference. An analysis of a four-way contingency table is presented for illustration of the techniques.

NONE

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Contingency Table Analysis						
Multi-Variate Data Analysis						
Manpower Data Analysis						

DD FORM 1 NOV 65 1473 (BACK)

(PAGE 2)

NONE

Security Classification

THE GEORGE WASHINGTON UNIVERSITY
Graduate School of Arts and Sciences
Econometric Research on Navy Manpower Problems

Abstract
of
Serial TR-1116
21 May 1973

THE ANALYSIS OF CONTINGENCY TABLES -
A METHODOLOGICAL EXPOSITION

by

S. Kullback

The analysis of the information in contingency tables is an aspect of multivariate (multiple variates) analysis with particular application to qualitative or categorical as well as quantitative variables.

The analysis is concerned with counts in multiway cross-classifications or multiple contingency tables. Multiway contingency tables, or cross-classifications of vectors of discrete random variables, provide a useful approach to the analysis of multivariate discrete data.

The method of analysis presented will bring out the various inter-relationships among the classificatory variables in a multiway cross-classification or contingency table in many dimensions. The illustration of the procedure is an application to Marine cohort data considering the relation of boot camp completion on home of record, level of education, and race.

The procedure is based on the Principle of Minimum Discrimination Information Estimation, associated statistics and Analyses of Information. General computer programs are available to provide the necessary results for inference. An analysis of a four-way contingency table is presented for illustration of the techniques.

TABLE OF CONTENTS

LIST OF TABLES

THE GEORGE WASHINGTON UNIVERSITY
Graduate School of Arts and Sciences
Econometric Research on Navy Manpower Problems

THE ANALYSIS OF CONTINGENCY TABLES - A METHODOLOGICAL EXPOSITION*

by

S. Kullback

1. Introduction

The primary purpose of this report is to present an exposition of the methodology underlying the analysis of the information in contingency tables. We shall stress the concepts, techniques, analyses and inferences without entering into extensive technical statistical proofs or detailed references to the bibliography at the end.

It is useful to note that we are concerned with an aspect of multivariate (multiple variates) analysis with particular application to qualitative or categorical as well as quantitative variables. The basic data we deal with are counts in multiway cross-classifications or multiple contingency tables. Multiway contingency tables, or cross-classifications of vectors of discrete random variables provide a useful approach to the analysis of multivariate discrete data.

As we shall see, the analytic procedures serve to bring out various interrelationships among the classificatory variables in a multiway cross-classification or contingency table in many dimensions. Classical problems in the historical development of the analysis of contingency

---

*This report was prepared under the Navy All Volunteer Force Manpower R&D Program of the Office of Naval Research under Contract Number N00014-67-A-0214, Task 0016, Project NR 347-024.

tables concerned themselves with such questions as the independence or conditional independence of the classificatory variables, or homogeneity or conditional homogeneity of the classificatory variables over time or space, for example. Such classical problems turn out to be special cases of the techniques we shall discuss. These techniques result in analyses which are essentially regression type analyses. As such they enable us to determine the relationship of one or more "dependent" qualitative or categorical variables of interest on a set of "independent" classificatory variables as well as the relative effects of changes in the "independent" variables on the "dependent" variables. In particular such problems as the determination of possible factors and measures of their effect in affecting failure or success in boot camp or decisions as to reenlistment lend themselves to the analysis we shall examine.

The methodology is based on the Principle of Minimum Discrimination Information Estimation, associated statistics and Analyses of Information. General computer programs are available to provide the data for the inferences.

2. Contingency Tables

We shall first present some examples of contingency tables to help clarify some of the terminology and, so to speak, set the scene. We shall use values obtained from the Marine COHORT File of 1966.

The simplest example of a contingency table is a one-way table with one classification, and several categories. The distribution of recruits by home of record is such an example, with four categories.

TABLE 2.1

HOME OF RECORD

East	North	West	South	Total
4201	4552	2840	5130	16723

There are not very many interesting questions that may arise for Table 2.1. The most likely question would be whether the distribution of

the occurrences is consistent with the distribution of potential recruits in the U.S. population by corresponding geographical classification.

A two-way contingency table arises when each observation has two classifications with different possible numbers of categories for each classification. An example of a 2x2 two-way contingency table arises when we distribute the recruits by Race and Success in Boot Camp.

TABLE 2.2a

Success in Boot Camp

		Fail	Pass	
Race	White	511	12637	13148
	Non-white	73	1629	1702
		584	14266	14850

We index the row categories by i, $i = 1$ White, $i = 2$ Non-white, and the column categories by j, $j = 1$ Fail, $j = 2$ Pass, and denote the occurrences by $x(ij)$, that is, the notation

Variable	Index	1	2
Race	i	White	Non-white
Boot Camp Completion	j	Fail	Pass

Thus Table 2.2a is represented as in Table 2.2b.

TABLE 2.2b

Success in Boot Camp

		Fail, $j=1$	Pass, $j=2$	
Race	White, $i=1$	$x(11)$	$x(12)$	$x(1\cdot)$
	Non-white, $i=2$	$x(21)$	$x(22)$	$x(2\cdot)$
		$x(\cdot 1)$	$x(\cdot 2)$	$x(\cdot\cdot) = n$

The sum of the entries across a row provide the corresponding row marginals and the sum of the entries down a column provide the corresponding column marginals. In the notation a dot is used to indicate summation

over a particular index. For Tables 2.2a and 2.2b the related values are

$$x(11) = 511$$
$$x(12) = 12637$$
$$x(21) = 73$$
$$x(22) = 1629$$
$$x(1\cdot) = x(11) + x(12) = 13148$$
$$x(2\cdot) = x(21) + x(22) = 1702$$
$$x(\cdot 1) = x(11) + x(21) = 584$$
$$x(\cdot 2) = x(12) + x(22) = 14266$$
$$x(\cdot\cdot) = x(11) + x(12) + x(21) + x(22) = 14850$$

but we usually use $n = x(\cdot\cdot)$.

For two-way 2x2 tables the primary question of interest is whether the row and column variables are independent. Thus in the two-way Table 2.2a the interest is in whether success in boot camp is the same for the two race categories. To answer this question one estimates the cell entries under the hypothesis of independence as a product of the marginals, that is, denoting the estimate by $x^*(ij)$ one uses $x^*(ij) = x(i\cdot)x(\cdot j)/n$. Some appropriate measure of the deviation between $x(ij)$ and $x^*(ij)$ is then used to determine whether the differences are "larger" than one would reasonably expect under the hypothesis of independence.

The estimated two-way table under the hypothesis or model of independence is given in Table 2.2c.

TABLE 2.2c

ESTIMATE UNDER INDEPENDENCE

$x^*(ij)$

	$j = 1$	$j = 2$	
$i = 1$	$x(1\cdot)x(\cdot 1)/n$	$x(1\cdot)x(\cdot 2)/n$	$x(1\cdot)$
$i = 2$	$x(2\cdot)x(\cdot 1)/n$	$x(2\cdot)x(\cdot 2)/n$	$x(2\cdot)$
	$x(\cdot 1)$	$x(\cdot 2)$	n

Note that the estimated table has the same marginals as the observed table $x(ij)$.

A common statistical measure of the association or interaction between the variables of a two-way 2x2 contingency table is the cross-product ratio, or its logarithm. The cross-product ratio is defined by

$$\frac{x(11)x(22)}{x(12)x(21)} , \tag{2.1}$$

though we shall be more concerned with its logarithm

$$\log \frac{x(11)x(22)}{x(12)x(21)} . \tag{2.2}$$

We shall use natural logarithms, that is, logarithms to the base e , rather than common logarithms to the base 10, because of the nature of the underlying mathematical statistical theory. Note that with the estimate for independence, or no association, the logarithm of the cross-product ratio is zero.

$$\log \frac{x^*(11)x^*(22)}{x^*(12)x^*(21)} = \log \frac{\frac{x(1\cdot)x(\cdot1)}{n}\,\frac{x(2\cdot)x(\cdot2)}{n}}{\frac{x(1\cdot)x(\cdot2)}{n}\,\frac{x(2\cdot)x(\cdot1)}{n}} = \log 1 = 0 . \tag{2.3}$$

The logarithm of the cross-product ratio is positive if the odds satisfy the inequalities

$$\frac{x(11)}{x(21)} > \frac{x(12)}{x(22)} \quad \text{or} \quad \frac{x(11)}{x(12)} > \frac{x(21)}{x(22)} ,$$

since then we get for the log-odds

$$\log \frac{x(11)x(22)}{x(12)x(21)} = \log \frac{x(11)}{x(21)} - \log \frac{x(12)}{x(22)} > 0$$

$$= \log \frac{x(11)}{x(12)} - \log \frac{x(21)}{x(22)} > 0 .$$

The logarithm of the cross-product ratio is negative if the odds satisfy the inequalities

$$\frac{x(11)}{x(21)} < \frac{x(12)}{x(22)} \quad \text{or} \quad \frac{x(11)}{x(12)} < \frac{x(21)}{x(22)} ,$$

since then we get for the log-odds

$$\log \frac{x(11)x(22)}{x(12)x(21)} = \log \frac{x(11)}{x(21)} - \log \frac{x(12)}{x(22)} < 0$$

$$= \log \frac{x(11)}{x(12)} - \log \frac{x(21)}{x(22)} < 0 .$$

The logarithm of the cross-product ratio thus varies from $-\infty$ to $+\infty$. Later we shall consider procedures for assessing the significance of the deviation of the logarithm of the cross-product ratio from zero, the value corresponding to no association or no interaction. Thus for the two-way Table 2.2a we have

$$\log \frac{511 \times 1629}{73 \times 12637} = \log \frac{832419}{922501} = \log 0.902$$

$$= -0.1031 .$$

We note that the odds of failure for White are 511/12637 = 0.04044 and the odds of failure for Non-white are 73/1629 = 0.04481.

Similar procedures apply to the case of a two-way rxc contingency table, that is, one with r rows and c columns.

TABLE 2.3a

TWO-WAY rxc CONTINGENCY TABLE

i \ j	1	2	...	c	
1	x(11)	x(12)	...	x(1c)	x(1·)
2	x(21)	x(22)	...	x(2c)	x(2·)
⋮	...	...	...	...	...
r	x(r1)	x(r2)	...	x(rc)	x(r·)
	x(·1)	x(·2)	...	x(·c)	n

Under a hypothesis or model of independence of row and column categories $x^*(ij) = x(i\cdot)x(\cdot j)/n$. Even if the row categories, say, are not randomly observed but selected with respect to some characteristic, say time or space, the mathematical procedures are still the same for determining whether the column categories are homogeneous over the row categories, time or space for instance. In the latter case we may consider the two-

way table as a set of one-way tables. Terms which cover both the case of independence and homogeneity are "association" or "interaction," that is, we question whether there is association or interaction among the variables.

The estimated two-way rxc contingency table under the hypothesis or model of independence is given in Table 2.3b.

TABLE 2.3b

ESTIMATE UNDER INDEPENDENCE

$x^*(ij)$

i \ j	1	2	...	c	
1	x(1•)x(•1)/n	x(1•)x(•2)/n	...	x(1•)x(•c)/n	x(1•)
2	x(2•)x(•1)/n	x(2•)x(•2)/n	...	x(2•)x(•c)/n	x(2•)
⋮	...	...	...	...	...
r	x(r•)x(•1)/n	x(r•)x(•2)/n	...	x(r•)x(•c)/n	x(r•)
	x(•1)	x(•2)	...	x(•c)	n

Note that the estimated table has the same marginals as the observed Table 2.3a.

A three-way contingency table arises when each observation has three classifications with different possible numbers of categories for each classification. The simplest three-way contingency table is 2x2x2, that is, with two categories for each classification. An example of a three-way 2x2x2 contingency table is the following cross-classification of recruits by AFQT (I and II, III and IV), Race (White, Non-white), Success in Boot Camp (Fail, Pass).

TABLE 2.4a

		AFQT				
		I and II		III and IV		
	Race	White	Non-white	White	Non-white	
BCC	Fail	143	0	618	130	891
	Pass	4989	113	7398	1459	13959
		5132	113	8016	1589	14850

We denote the occurrences in the three-way Table 2.4a by $x(ijk)$ with the notation

Variable	Index	1	2
AFQT	i	I and II	III and IV
Race	j	White	Non-white
Boot Camp Completion	k	Fail	Pass

In the general notation we have Table 2.4b.

TABLE 2.4b

	i = 1		i = 2		
	j = 1	j = 2	j = 1	j = 2	
k = 1	x(111)	x(121)	x(211)	x(221)	x(••1)
k = 2	x(112)	x(122)	x(212)	x(222)	x(••2)
	x(11•)	x(12•)	x(21•)	x(22•)	n

The two-way marginals are

$$x(11\cdot) = x(111) + x(112)$$
$$x(12\cdot) = x(121) + x(122)$$
$$x(21\cdot) = x(211) + x(212)$$
$$x(22\cdot) = x(221) + x(222)$$
$$x(1\cdot 1) = x(111) + x(121)$$
$$x(1\cdot 2) = x(112) + x(122)$$
$$x(2\cdot 1) = x(211) + x(221)$$
$$x(2\cdot 2) = x(212) + x(222)$$

$$x(\cdot 11) = x(111) + x(211)$$
$$x(\cdot 12) = x(112) + x(212)$$
$$x(\cdot 21) = x(121) + x(221)$$
$$x(\cdot 22) = x(122) + x(222) \; .$$

The one-way marginals are

$$x(1\cdot\cdot) = x(111) + x(112) + x(121) + x(122) = x(11\cdot) + x(12\cdot)$$
$$x(2\cdot\cdot) = x(211) + x(212) + x(221) + x(222) = x(21\cdot) + x(22\cdot)$$
$$x(\cdot 1\cdot) = x(111) + x(112) + x(211) + x(212) = x(11\cdot) + x(21\cdot)$$
$$x(\cdot 2\cdot) = x(121) + x(122) + x(221) + x(222) = x(12\cdot) + x(22\cdot)$$
$$x(\cdot\cdot 1) = x(111) + x(121) + x(211) + x(221) = x(1\cdot 1) + x(2\cdot 1)$$
$$x(\cdot\cdot 2) = x(112) + x(122) + x(212) + x(222) = x(1\cdot 2) + x(2\cdot 2)$$

The entries $x(ijk)$ in Table 2.4b may also be considered as three-way marginals.

With more variables there are more possible questions of interest. One may be interested in whether any pair of the variables are independent or show no interaction or association. One may be interested in conditional independence, that is, whether a pair of variables are independent given the third variable. One may be interested in whether the three variables are mutually independent or whether one of the variables is independent of the pair of the other variables. These questions of independence, no interaction or association are all answered by considering estimates which are explicitly represented in terms of products of various marginals. We list some of these estimates.

Mutual independence of i, j, and k	$x_1^*(ijk) = x(i\cdot\cdot)x(\cdot j\cdot)x(\cdot\cdot k)/n^2$
Independence of i and (jk) jointly	$x_a^*(ijk) = x(i\cdot\cdot)x(\cdot jk)/n$
Conditional independence of i and j given k	$x_b^*(ijk) = x(i\cdot k)x(\cdot jk)/x(\cdot\cdot k)$

As might be expected, these estimates also apply in the general three-way rxsxt contingency table.

We note that the estimate under mutual independence of i, j, and k has the same one-way marginals as the observed table $x(ijk)$.

$$x_1^*(111) = x(1\cdot\cdot)x(\cdot 1\cdot)x(\cdot\cdot 1)/n^2$$

$$x_1^*(112) = x(1\cdot\cdot)x(\cdot 1\cdot)x(\cdot\cdot 2)/n^2$$

$$x_1^*(121) = x(1\cdot\cdot)x(\cdot 2\cdot)x(\cdot\cdot 1)/n^2$$

$$x_1^*(122) = x(1\cdot\cdot)x(\cdot 2\cdot)x(\cdot\cdot 2)/n^2$$

$$x_1^*(211) = x(2\cdot\cdot)x(\cdot 1\cdot)x(\cdot\cdot 1)/n^2$$

$$x_1^*(212) = x(2\cdot\cdot)x(\cdot 1\cdot)x(\cdot\cdot 2)/n^2$$

$$x_1^*(221) = x(2\cdot\cdot)x(\cdot 2\cdot)x(\cdot\cdot 1)/n^2$$

$$x_1^*(222) = x(2\cdot\cdot)x(\cdot 2\cdot)x(\cdot\cdot 2)/n^2$$

$$\begin{aligned} x_1^*(1\cdot\cdot) &= x_1^*(111) + x_1^*(112) + x_1^*(121) + x_1^*(122) \\ &= x(1\cdot\cdot)x(\cdot 1\cdot)/n + x(1\cdot\cdot)x(\cdot 2\cdot)/n \\ &= x(1\cdot\cdot) \end{aligned}$$

$$\begin{aligned} x_1^*(2\cdot\cdot) &= x_1^*(211) + x_1^*(212) + x_1^*(221) + x_1^*(222) \\ &= x(2\cdot\cdot)x(\cdot 1\cdot)/n + x(2\cdot\cdot)x(\cdot 2\cdot)/n \\ &= x(2\cdot\cdot) \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot 1\cdot) &= x_1^*(111) + x_1^*(112) + x_1^*(211) + x_1^*(212) \\ &= x(1\cdot\cdot)x(\cdot 1\cdot)/n + x(2\cdot\cdot)x(\cdot 1\cdot)/n \\ &= x(\cdot 1\cdot) \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot 2\cdot) &= x_1^*(121) + x_1^*(122) + x_1^*(221) + x_1^*(222) \\ &= x(\cdot 2\cdot) \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot\cdot 1) &= x_1^*(111) + x_1^*(121) + x_1^*(211) + x_1^*(221) \\ &= x(\cdot\cdot 1) \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot\cdot 2) &= x_1^*(112) + x_1^*(122) + x_1^*(212) + x_1^*(222) \\ &= x(\cdot\cdot 2) \end{aligned}$$

However, the two-way marginals of the estimate under mutual independence of i, j, and k differ from the two-way marginals of the observed table $x(ijk)$. Thus, for example,

$$x_1^*(11\cdot) = x_1^*(111) + x_1^*(112)$$
$$= x(1\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot1)/n^2 + x(1\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot2)/n^2$$
$$= x(1\cdot\cdot)x(\cdot1\cdot)/n \ ,$$

and the latter value is not necessarily equal to $x(11\cdot)$.

The estimate under the hypothesis or model of independence of i and (jk) jointly has the same one-way marginals and the same two-way jk-marginal as the observed table $x(ijk)$.

$$x_a^*(111) = x(1\cdot\cdot)x(\cdot11)/n$$
$$x_a^*(112) = x(1\cdot\cdot)x(\cdot12)/n$$
$$x_a^*(121) = x(1\cdot\cdot)x(\cdot21)/n$$
$$x_a^*(122) = x(1\cdot\cdot)x(\cdot22)/n$$
$$x_a^*(211) = x(2\cdot\cdot)x(\cdot11)/n$$
$$x_a^*(212) = x(2\cdot\cdot)x(\cdot12)/n$$
$$x_a^*(221) = x(2\cdot\cdot)x(\cdot21)/n$$
$$x_a^*(222) = x(2\cdot\cdot)x(\cdot22)/n$$
$$x_a^*(1\cdot\cdot) = x_a^*(111) + x_a^*(112) + x_a^*(121) + x_a^*(122)$$
$$= x(1\cdot\cdot)x(\cdot11)/n + x(1\cdot\cdot)x(\cdot12)/n + x(1\cdot\cdot)x(\cdot21)/n + x(1\cdot\cdot)x(\cdot22)/n$$
$$= x(1\cdot\cdot)[x(\cdot11) + x(\cdot12) + x(\cdot21) + x(\cdot22)]/n$$
$$= x(1\cdot\cdot)$$

Similar results follow for the other one-way marginals.

$$x_a^*(\cdot11) = x_a^*(111) + x_a^*(211)$$
$$= x(1\cdot\cdot)x(\cdot11)/n + x(2\cdot\cdot)x(\cdot11)/n$$
$$= x(\cdot11)$$
$$x_a^*(\cdot12) = x_a^*(112) + x_a^*(212)$$
$$= x(1\cdot\cdot)x(\cdot12)/n + x(2\cdot\cdot)x(\cdot12)/n$$
$$= x(\cdot12)$$

$$x_a^*(\cdot 21) = x_a^*(121) + x_a^*(221)$$
$$= x(1\cdot\cdot)x(\cdot 21)/n + x(2\cdot\cdot)x(\cdot 21)/n$$
$$= x(\cdot 21)$$
$$x_a^*(\cdot 22) = x_a^*(122) + x_a^*(222)$$
$$= x(1\cdot\cdot)x(\cdot 22)/n + x(2\cdot\cdot)x(\cdot 22)/n$$
$$= x(\cdot 22)$$

However, for the other two-way marginals, for example,

$$x_a^*(11\cdot) = x_a^*(111) + x_a^*(112)$$
$$= x(1\cdot\cdot)x(\cdot 11)/n + x(1\cdot\cdot)x(\cdot 12)/n$$
$$= x(1\cdot\cdot)[x(\cdot 11) + x(\cdot 12)]/n$$
$$= x(1\cdot\cdot)x(\cdot 1\cdot)/n ,$$

and the latter value is not necessarily equal to $x(11\cdot)$.

$$x_a^*(1\cdot 1) = x_a^*(111) + x_a^*(121)$$
$$= x(1\cdot\cdot)x(\cdot 11)/n + x(1\cdot\cdot)x(\cdot 21)/n$$
$$= x(1\cdot\cdot)[x(\cdot 11) + x(\cdot 21)]/n$$
$$= x(1\cdot\cdot)x(\cdot\cdot 1)/n ,$$

and the latter value is not necessarily equal to $x(1\cdot 1)$.

The estimate under the hypothesis or model of conditional independence of i and j given k has the same one-way marginals and the same two-way ik- and jk-marginals as the observed table $x(ijk)$.

$$x_b^*(111) = x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1)$$
$$x_b^*(112) = x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2)$$
$$x_b^*(121) = x(1\cdot 1)x(\cdot 21)/x(\cdot\cdot 1)$$
$$x_b^*(122) = x(1\cdot 2)x(\cdot 22)/x(\cdot\cdot 2)$$
$$x_b^*(211) = x(2\cdot 1)x(\cdot 11)/x(\cdot\cdot 1)$$

$$x_b^*(212) = x(2\cdot 2)x(\cdot 12)/x(\cdot\cdot 2)$$

$$x_b^*(221) = x(2\cdot 1)x(\cdot 21)/x(\cdot\cdot 1)$$

$$x_b^*(222) = x(2\cdot 2)x(\cdot 22)/x(\cdot\cdot 2)$$

$$\begin{aligned} x_b^*(1\cdot\cdot) &= x_b^*(111) + x_b^*(112) + x_b^*(121) + x_b^*(122) \\ &= x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) + x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) \\ &\quad + x(1\cdot 1)x(\cdot 21)/x(\cdot\cdot 1) + x(1\cdot 2)x(\cdot 22)/x(\cdot\cdot 2) \\ &= x(1\cdot 1) + x(1\cdot 2) = x(1\cdot\cdot) \ . \end{aligned}$$

Similar results follow for the other one-way marginals.

$$\begin{aligned} x_b^*(1\cdot 1) &= x_b^*(111) + x_b^*(121) \\ &= x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) + x(1\cdot 1)x(\cdot 21)/x(\cdot\cdot 1) \\ &= x(1\cdot 1) \\ x_b^*(1\cdot 2) &= x_b^*(112) + x_b^*(122) \\ &= x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) + x(1\cdot 2)x(\cdot 22)/x(\cdot\cdot 2) \\ &= x(1\cdot 2) \ , \end{aligned}$$

and in a similar manner we have

$$x_b^*(2\cdot 1) = x(2\cdot 1) \ , \ x_b^*(2\cdot 2) = x(2\cdot 2)$$

$$\begin{aligned} x_b^*(\cdot 11) &= x_b^*(111) + x_b^*(211) \\ &= x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) + x(2\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) \\ &= x(\cdot 11) \\ x_b^*(\cdot 12) &= x_b^*(112) + x_b^*(212) \\ &= x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) + x(2\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) \\ &= x(\cdot 12) \ , \end{aligned}$$

and in a similar manner we have

$$x_b^*(\cdot 21) = x(\cdot 21) \ , \ x_b^*(\cdot 22) = x(\cdot 22) \ .$$

However, for the other two-way marginals

$$x_b^*(11\cdot) = x_b^*(111) + x_b^*(112)$$
$$= x(1\cdot1)x(\cdot11)/x(\cdot\cdot1) + x(1\cdot2)x(\cdot12)/x(\cdot\cdot2) \ ,$$

and the latter value is not necessarily equal to $x(11\cdot)$.

We remark that one of the constraints in the determination of the estimates was that they have certain marginals the same as the observed table.

For the three-way table in addition to the types of independence, interaction or association just discussed, there arises an additional one, important historically and practically. This is known as no three-factor or no second-order interaction. No three-factor or no second-order interaction implies that the logarithm of the association measured by the cross-product ratio for any two of the variables is the same for all the values of the third variable, that is, there is no second-order interaction if

$$(2.4) \qquad \begin{cases} \ln \dfrac{x(111)x(221)}{x(121)x(211)} = \ln \dfrac{x(112)x(222)}{x(122)x(212)} \ , & i,\ j \\ \ln \dfrac{x(111)x(212)}{x(112)x(211)} = \ln \dfrac{x(121)x(222)}{x(122)x(221)} \ , & i,\ k \\ \ln \dfrac{x(111)x(122)}{x(112)x(121)} = \ln \dfrac{x(211)x(222)}{x(212)x(221)} \ , & j,\ k \ . \end{cases}$$

One is concerned with the possible hypothesis or model of no second-order interaction when none of the other types of independence are found. However, in this case, the corresponding estimate cannot be expressed explicitly in terms of observed marginals although the estimate is constrained to have the same two-way marginals as the observed table. Straightforward iterative procedures exist to determine the estimate under the hypothesis or model of no second-order interaction. For the general three-way rxsxt contingency table there are of course many more relations among the log cross-product ratios like (2.4) which must be satisfied, but the iterative procedures to determine the estimate extend to the general case with no difficulty.

We may be concerned with a set of two-way tables for which it is of interest to determine whether they are homogeneous with respect to a

third factor, say space or time. Such problems may also be treated as three-way contingency tables using the space or time factor as the third classification.

For four-way and higher order contingency tables the problem of presentation of the data increases, as do the variety and number of questions about relationships of possible interest and varieties of interaction. The basic ideas, concepts, notation and terminology we have discussed for the one-, two- and three-way contingency tables extend to the more general cases as we consider the methodology.

3. Discrimination Information

To make the discussion more specific and with no essential restriction on the generality, we shall present it in terms of the analysis of four-way contingency tables. Let us consider the collection of four-way contingency tables RxSxTxU of dimension rxsxtxu . For convenience let us denote the aggregate of all cell identifications by Ω with individual cells identified by ω so that the generic variable is $\omega = (i,j,k,\ell)$, $i = 1,\ldots,r$, $j = 1,\ldots,s$, $k = 1,\ldots,t$, $\ell = 1,\ldots,u$. Suppose there are two probability distributions or contingency tables (we shall use these terms interchangeably) defined over the space Ω , say $p(\omega)$, $\pi(\omega)$, $\sum_{\Omega} p(\omega) = 1$, $\sum_{\Omega} \pi(\omega) = 1$. The discrimination information is defined by

$$(3.1) \qquad I(p:\pi) = \sum_{\Omega} p(\omega) \,\ell n \frac{p(\omega)}{\pi(\omega)} \quad .$$

The basis for this definition, its properties, and relation to other definitions of information measures will not be considered in detail in this exposition. For the particular types of application to which we shall restrict this exposition the π-distribution, $\pi(\omega)$, in the definition (3.1) according to the problem of interest may either be specified, or it may be an estimated distribution. The p-distribution, $p(\omega)$, in the definition (3.1) ranges over or is a member of a family of distributions of interest.

Of the various properties of $I(p:\pi)$ we mention in particular the fact that $I(p:\pi) > 0$ and $= 0$ if and only if $p(\omega) = \pi(\omega)$.

4. Minimum Discrimination Information Estimation

Many problems in the analysis of contingency tables may be characterized as estimating a distribution or contingency table subject to certain restraints and then comparing the estimated table with an observed table to determine whether the observed table satisfies a null hypothesis or model implied by the restraints. In accordance with the principle of minimum discrimination information estimation we determine that member of the collection or family of p-distributions satisfying the restraints which minimizes the discrimination information $I(p:\pi)$ over all members of the family of pertinent p-distributions. We denote the minimum discrimination information estimate by $p^*(\omega)$ so that

$$(4.1) \qquad I(p^*:\pi) = \Sigma\, p^*(\omega)\, \ln \frac{p^*(\omega)}{\pi(\omega)} = \min I(p:\pi)\ .$$

Unless otherwise stated, the summation is over Ω which will be omitted.

In a wide class of problems which can be characterized as "smoothing" or fitting an observed contingency table the restraints specify that the estimated distribution or contingency table have some set of marginals which are the same as those of an observed contingency table. In such cases $\pi(\omega)$ is taken to be either the uniform distribution $\pi(ijk\ell) = 1/rstu$ or a distribution already estimated subject to restraints contained in and implied by the restraints under examination. The latter case includes the classical hypotheses of independence, conditional independence, homogeneity, conditional homogeneity and interaction, all of which can be considered as instances of generalized independence and will be considered in some detail in this report. By generalized independence is meant the fact that the estimates may be expressed as a product of factors which are functions of appropriate marginals.

5. Minimum Discrimination Information Statistic

To test whether an observed contingency table is consistent with the null hypothesis or model as represented by the minimum discrimination information estimate we compute a measure of the deviation between the observed distribution and the appropriate estimate by the minimum discrimination information statistic. For notational convenience and later

computational convenience let us denote the estimated contingency table in terms of occurrences by $x^*(\omega) = np^*(\omega)$. For the "smoothing" or fitting class of problems, that is, with the restraints implied by a set of observed marginals (those of a generalized independence hypothesis), the minimum discrimination information (m.d.i.) statistic is

$$2I(x:x^*) = 2\Sigma\ x(\omega)\ \ln \frac{x(\omega)}{x^*(\omega)} \tag{5 1}$$

which is asymptotically distributed as a χ^2 with appropriate degrees of freedom under the null hypothesis.

The statistic in (5.1) is also minus twice the logarithm of the classic likelihood ratio statistic but this is not necessarily true for other kinds of applications of the general theory.

6. Minimum Discrimination Information Theorem

We now present a theorem which is the basis for the principle of minimum discrimination information estimation and its applications. We shall present it in a form related to the context of this discussion on the analysis of contingency tables.

Let us consider the space Ω mentioned in Section 3 and the discrimination information introduced in (3.1). Suppose now, for example, that there are three linearly independent statistics of interest defined over the space Ω ,

$$T_1(\omega)\ ,\ T_2(\omega)\ ,\ T_3(\omega)\ . \tag{6.1}$$

Let us determine the value of $p(\omega)$ which minimizes the discrimination information

$$I(p:\pi) = \Sigma\ p(\omega)\ \ln \frac{p(\omega)}{\pi(\omega)} \tag{6.2}$$

over the family of p-distributions which satisfies the restraints

$$(6.3)\qquad \begin{aligned} \Sigma\, T_1(\omega)p(\omega) &= \theta_1^* \\ \Sigma\, T_2(\omega)p(\omega) &= \theta_2^* \\ \Sigma\, T_3(\omega)p(\omega) &= \theta_3^* \end{aligned}$$

where θ_1^* , θ_2^* , θ_3^* are specified values, and $\pi(\omega)$ is a fixed distribution.

If $\pi(\omega)$ satisfies the restraints (6.3), then of course the minimum value of $I(p:\pi)$ is zero and the minimizing distribution is $p^*(\omega) = \pi(\omega)$. More generally, the minimum discrimination information theorem states that the minimizing distribution is given by

$$(6.4)\qquad p^*(\omega) = \frac{\exp\,(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))\pi(\omega)}{M(\tau_1,\tau_2,\tau_3)}$$

where

$$(6.5)\qquad M(\tau_1,\tau_2,\tau_3) = \Sigma\, \exp\,(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))\pi(\omega)$$

is a normalizing factor so that $\Sigma\, p^*(\omega) = 1$, and the τ's are parameters which technically are in essence undetermined Lagrange multipliers whose values are defined in terms of θ_1^* , θ_2^* , θ_3^* by

$$\begin{aligned}
\theta_1^* &= \frac{\partial}{\partial\tau_1} \ln M(\tau_1,\tau_2,\tau_3) \\
&= (\Sigma\, \exp\,(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))T_1(\omega)\pi(\omega))/M(\tau_1,\tau_2,\tau_3) \\
&= \Sigma\, T_1(\omega)p^*(\omega) \\
\theta_2^* &= \frac{\partial}{\partial\tau_2} \ln M(\tau_1,\tau_2,\tau_3) \\
&= (\Sigma\, \exp\,(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))T_2(\omega)\pi(\omega))/M(\tau_1,\tau_2,\tau_3) \qquad (6.6) \\
&= \Sigma\, T_2(\omega)p^*(\omega) \\
\theta_3^* &= \frac{\partial}{\partial\tau_3} \ln M(\tau_1,\tau_2,\tau_3) \\
&= (\Sigma\, \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))T_3(\omega)\pi(\omega))/M(\tau_1,\tau_2,\tau_3) \\
&= \Sigma\, T_3(\omega)p^*(\omega)\ .
\end{aligned}$$

We can now state a number of consequences of the preceding.

We note first that $p^*(\omega)$ is a member of an exponential family of distributions generated by $\pi(\omega)$ and as such has the desirable statistical properties of members of an exponential family which include all the common and classic distributions. We may also write (6.4) as

$$(6.7) \qquad \ln \frac{p^*(\omega)}{\pi(\omega)} = -\ln M(\tau_1,\tau_2,\tau_3) + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

$$= L + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

with $L = -\ln M(\tau_1,\tau_2,\tau_3)$. The regression or log-linear expression in (6.7) for $\ln (p^*(\omega)/\pi(\omega))$ with $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ as the explanatory variables and τ_1 , τ_2 , τ_3 as the regression coefficients plays an important role in the analysis we shall consider.

We note next that the minimum value of the discrimination information (6.2) is

$$(6.8) \qquad I(p^*:\pi) = \tau_1\theta_1^* + \tau_2\theta_2^* + \tau_3\theta_3^* - \ln M(\tau_1,\tau_2,\tau_3)$$

where the θ^*'s are defined in (6.3) and the τ's are determined to satisfy (6.6). Using the value in (6.7) it may be shown that if $p(\omega)$ is any member of the family of distributions satisfying (6.3), then

$$(6.9) \qquad I(p:\pi) = I(p^*:\pi) + I(p:p^*) \ .$$

The pythagorean type property (6.9) plays an important role in the analysis of information tables.

7. Computational Procedures

An "experiment" has been designed and observations made resulting in a multi-dimensional contingency table with the desired classifications and categories. All the information the analyst hopes to obtain from the "experiment" is contained in the contingency table. In the process of analysis, the aim is to fit the observed table by a minimal or parsimonious number of parameters depending on some or all of the marginals, that is,

to find out how much of this total information is contained in a summary consisting of sets of marginals. Indeed, the relationship between the concept of independence or association and interaction in contingency tables and the role the marginals play is evidenced in the historical developments in the extensive literature on the analysis of contingency tables. Thus, the θ^*'s in the preceding discussion will be the marginals of interest.

7.1. The $T(\omega)$ Functions. The $T(\omega)$ functions for the RxSxTxU table turn out to be a basic set of simple functions and their various products. Thus, for example, the $T(\omega)$ function associated with the one-way marginal $p(2...)$ is

$$(7.1) \qquad T_2^R(ijk\ell) = 1 \quad \text{for } i = 2 \text{ , any } j,k,\ell$$
$$= 0 \quad \text{otherwise}$$

since

$$(7.2) \qquad \Sigma\, p(ijk\ell)\, T_2^R(ijk\ell) = p(2...) \ .$$

Similarly the $T(\omega)$ function associated with the one-way marginal $p(..3.)$, for example, is

$$(7.3) \qquad T_3^T(ijk\ell) = 1 \quad \text{for } k = 3 \text{ , any } i,j,\ell$$
$$= 0 \quad \text{otherwise}$$

since

$$(7.4) \qquad \Sigma\, p(ijk\ell)\, T_3^T(ijk\ell) = p(..3.) \ .$$

Thus for the rxsxtxu table there are

$$(7.5) \qquad \begin{array}{l} (r-1) \text{ linearly independent functions } T_\alpha^R(ijk\ell),\ \alpha = 1,\ldots,r-1 \\ (s-1) \text{ linearly independent functions } T_\beta^S(ijk\ell),\ \beta = 1,\ldots,s-1 \\ (t-1) \text{ linearly independent functions } T_\gamma^T(ijk\ell),\ \gamma = 1,\ldots,t-1 \\ (u-1) \text{ linearly independent functions } T_\delta^U(ijk\ell),\ \delta = 1,\ldots,u-1 \ , \end{array}$$

since, for example,

$$\sum_{\alpha=1}^{r} \Sigma\, T_\alpha^R(ijk\ell) = rstu \ .$$

We have arbitrarily excluded the functions corresponding to $\alpha = r$, $\beta = s$, $\gamma = t$, $\delta = u$ as a matter of convenience. We could have selected $\alpha = 1, \beta = 1, \gamma = 1, \delta = 1$ or any other set of values.

The $T(\omega)$ function associated with the two-way marginal $p(12..)$ say, is $T_1^R(ijk\ell)\ T_2^S(ijk\ell)$ since from the definition of $T_1^R(ijk\ell)$ and $T_2^S(ijk\ell)$ it may be seen that

$$(7.6) \qquad T_1^R(ijk\ell)\ T_2^S(ijk\ell) = 1 \quad \text{for } i = 1,\ j = 2, \text{ any } k,\ell$$
$$= 0 \quad \text{otherwise}$$

and

$$(7.7) \qquad \Sigma\ p(ijk\ell)\ T_1^R(ijk\ell)\ T_2^S(ijk\ell) = p(12..)\ .$$

For convenience we shall write $T_\alpha^R(ijk\ell)\ T_\beta^S(ijk\ell) = T_{\alpha\beta}^{RS}(ijk\ell)$, etc. Thus the $T(\omega)$ function associated with any two-way marginal is a product of two appropriate functions of the set (7.5).

Similarly the $T(\omega)$ function associated with any three-way marginal will be a product of three of the appropriate functions of the set (7.5), for example,

$$(7.8) \qquad \Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^T(ijk\ell)\ T_3^U(ijk\ell) = p(2.13)\ .$$

For convenience we shall write $T_\alpha^R(ijk\ell)\ T_\beta^S(ijk\ell)\ T_\gamma^T(ijk\ell) = T_{\alpha\beta\gamma}^{RST}(ijk\ell)$, etc.

Similarly the $T(\omega)$ function associated with any four-way marginal will be a product of four of the appropriate functions of the set (7.5), for example,

$$(7.9) \qquad \Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^S(ijk\ell)\ T_1^T(ijk\ell)\ T_2^U(ijk\ell) = p(2112)\ .$$

For convenience we shall write $T_\alpha^R(ijk\ell)\ T_\beta^S(ijk\ell)\ T_\gamma^T(ijk\ell)\ T_\delta^U(ijk\ell) = T_{\alpha\beta\gamma\delta}^{RSTU}(ijk\ell)$.

We note that there are a total of

$$(7.10)\quad \begin{cases} N_1 = (r-1) + (s-1) + (t-1) + (u-1) \\ N_2 = (r-1)(s-1) + (r-1)(t-1) + (r-1)(u-1) + (s-1)(t-1) \\ \qquad + (s-1)(u-1) + (t-1)(u-1) \\ N_3 = (r-1)(s-1)(t-1) + (r-1)(s-1)(u-1) + (r-1)(t-1)(u-1) \\ \qquad + (s-1)(t-1)(u-1) \\ N_4 = (r-1)(s-1)(t-1)(u-1) \ , \end{cases}$$

respectively, of the simple linearly independent functions and their products two, three, four at a time. It may be verified that

$$(7.11)\qquad rstu - 1 = N = N_1 + N_2 + N_3 + N_4 \ .$$

These values of the number of $T(\omega)$ functions (or associated tau parameters) appear as appropriate degrees of freedom in the analysis of information tables.

7.2. <u>The Estimated $p^*(\omega)$ Values</u>. In the usual least squares regression analysis procedure, one first computes the regression coefficients and then gets the values of the estimates. In the methodology we use we reverse the procedure. Instead of trying to obtain the values of the τ's from (6.6) (which is possible) we shall first obtain the values of the estimates $p^*(\omega)$ by a straightforward convergent iterative procedure and then derive the values of the τ's from (6.7). We shall not discuss the details of the iteration here, as they are in the computer program and have been described elsewhere. The iteration may be described as successively cycling through adjustments of the marginals of interest starting with the $\pi(\omega)$ distribution until a desired accuracy of agreement between the set of observed marginals of interest and the computed marginals has been attained.

7.3. <u>The τ Values or Interaction Parameters</u>. From the definitions of the $T(\omega)$ functions in Section 7.1 it is clear that they take on only the values 0 or 1 for each value of ω . From the nature of the $T(\omega)$

functions the set of regression or log-linear Equations (6.7) will have some with a single τ value which can be determined. Then there will be a set with one additional unknown value and some of the τ's already determined. These new unknown τ values can be then determined. This process of successive evaluation is carried on until all the values of τ are determined. They are also available as output of a general computer program.

8. Graphic Representation

A useful graphic representation of the log-linear regression (6.7) is given in Figure 8.1 for a 2x2x2x2 contingency table. This is the analogue of the design matrix in normal regression theory. The blank spaces in Figure 8.1 represent zero values. The $(ijk\ell)$-columns are the cell identifications in the same lexographic order as the cell entries for the estimates in the computer output. Column 1 corresponds to L which is essentially a normalizing factor. Each of the columns 2 to 16 represents the corresponding values of the $T(\omega)$ functions, columns 2 to 5 those for the one-way marginals, columns 6 to 11 those for the two-way marginals, columns 12 to 15 those for the three-way marginals, and column 16 that for the four-way marginal. For convenience the columns are also arranged in lexographic order. The tau parameter associated with the $T(\omega)$ function is given at the head of the column. The full representation with all the columns of Figure 8.1 generates the observed values. Thus the rows represent

$$(8.1) \qquad \ln \frac{p(ijk\ell)}{\pi(ijk\ell)} = \ln \frac{x(ijk\ell)}{n\pi(ijk\ell)} = L + \tau_1^i T_1^i(ijk\ell) + \dots + \tau_{11}^{ij} T_{11}^{ij}(ijk\ell) + \dots + \tau_{111}^{ijk} T_{111}^{ijk}(ijk\ell) + \dots + \tau_{1111}^{ijk\ell} T_{1111}^{ijk\ell}(ijk\ell)$$

where $\pi(ijk\ell)$ in the 2x2x2x2 case is 1/2x2x2x2 and the numerical values of L and the taus depend on the observed values $x(ijk\ell)$. The design matrix corresponding to an estimate uses only those columns associated with the marginals explicit and implied in the fitting process. This is a reflection of the fact that higher order marginals imply certain

ω	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
i j k ℓ	L	τ_1^i	τ_1^j	τ_1^k	τ_1^ℓ	τ_{11}^{ij}	τ_{11}^{ik}	$\tau_{11}^{i\ell}$	τ_{11}^{jk}	$\tau_{11}^{j\ell}$	$\tau_{11}^{k\ell}$	τ_{111}^{ijk}	$\tau_{111}^{ij\ell}$	$\tau_{111}^{ik\ell}$	$\tau_{111}^{jk\ell}$	$\tau_{1111}^{ijk\ell}$
1 1 1 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1 1 1 2	1	1	1	1		1	1		1			1				
1 1 2 1	1	1	1		1	1		1		1			1			
1 1 2 2	1	1	1			1										
1 2 1 1	1	1		1	1		1	1			1			1		
1 2 1 2	1	1		1			1									
1 2 2 1	1	1			1			1								
1 2 2 2	1	1														
2 1 1 1	1		1	1	1				1	1	1				1	
2 1 1 2	1		1	1					1							
2 1 2 1	1		1		1					1						
2 1 2 2	1		1													
2 2 1 1	1			1	1						1					
2 2 1 2	1			1												
2 2 2 1	1				1											
2 2 2 2	1															

Figure 8.1. Graphic representation.

lower order marginals, for example, the two-way marginal $x(ij..)$ implies, by summation over i and j, the one-way marginals $x(.j..)$, $x(i...)$, and the total $n = x(....)$. Thus the estimate based on fitting the one-way marginals will use only columns 1-5. The values of L and the taus for this estimate will be different from those for $x(ijk\ell)$ and depend on the estimate $x_1^*(ijk\ell)$. Thus if we denote the estimate based on fitting the one-way marginals as $x_1^*(ijk\ell)$, the representation in Figure 8.1 implies

$$
\left\{
\begin{aligned}
\ell n \frac{x_1^*(1111)}{n\pi} &= L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_1^\ell \\
\ell n \frac{x_1^*(1112)}{n\pi} &= L + \tau_1^i + \tau_1^j + \tau_1^k \\
\vdots \qquad &\quad \vdots \qquad\quad \vdots \\
\ell n \frac{x_1^*(2222)}{n\pi} &= L \,.
\end{aligned}
\right.
\tag{8.2}
$$

The estimate based on fitting the two-way marginals will use columns 1-11 since the two-way marginals also imply the one-way marginals. The values of L and the taus for this estimate will be different from those for the observed values or other estimates and depend on the values of the estimate which we denote by $x_2^*(ijk\ell)$. For the estimate fitting the two-way marginals the representation in Figure 8.1 implies

$$(8.3)\quad \begin{cases} \ln \dfrac{x_2^*(1111)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_1^\ell + \tau_{11}^{ij} + \tau_{11}^{ik} + \tau_{11}^{i\ell} + \tau_{11}^{jk} + \tau_{11}^{j\ell} + \tau_{11}^{k\ell} \\ \ln \dfrac{x_2^*(1112)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_{11}^{ij} + \tau_{11}^{ik} + \tau_{11}^{jk} \\ \quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \ln \dfrac{x_2^*(2222)}{n\pi} = L . \end{cases}$$

The representation for the uniform distribution corresponds to column 1 only.

Note that in accordance with (7.10) and (7.11)

$$\begin{aligned} N_1 &= 1 + 1 + 1 + 1 && = 4 \text{ (columns 2 to 5)} \\ N_2 &= 1 + 1 + 1 + 1 + 1 + 1 && = 6 \text{ (columns 6 to 11)} \\ N_3 &= 1 + 1 + 1 + 1 && = 4 \text{ (columns 12 to 15)} \\ N_4 &= 1 && = 1 \text{ (column 16)} \\ N &= 16 - 1 = 15 && = 4 + 6 + 4 + 1 . \end{aligned}$$

9. Analysis of Information

Although the preceding discussion has at times been in terms of probabilities, estimated probabilities or relative frequencies, in practice it has been found more convenient not to divide everything by n , the total number of occurrences, and deal with observed or estimated occurrences, that is, with $n\pi(ijk\ell) = n/rstu$, $x(ijk\ell)$, $x(i...)$, $x(.jk.)$, $x^*(ijk\ell) = np^*(ijk\ell)$, etc. The analysis of information is based on the fundamental relation (6.9) for the minimum discrimination information statistics. Specifically if $np_a^*(\omega) = x_a^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_a of given marginals and $x_b^*(\omega)$ is the

minimum discrimination information estimate corresponding to a set H_b of given marginals, where H_a is explicitly or implicitly contained in H_b, then the basic relations are

$$
\begin{aligned}
2I(x:n\pi) &= 2I(x_a^*:n\pi) + 2I(x:x_a^*) \\
2I(x:n\pi) &= 2I(x_b^*:n\pi) + 2I(x:x_b^*) \\
2I(x_b^*:n\pi) &= 2I(x_a^*:n\pi) + 2I(x_b^*:x_a^*) \\
2I(x:x_a^*) &= 2I(x_b^*:x_a^*) + 2I(x:x_b^*)
\end{aligned}
\tag{9.1}
$$

with a corresponding additive relation for the associated degrees of freedom.

In terms of the representation in (6.4) or (6.7) or Figure 8.1 as an exponential family, for our discussion, the two extreme cases are the uniform distribution for which all τ's are zero, and the observed contingency table or distribution for which all $N = rstu - 1$ τ's are needed.

Measures of the form $2I(x:x_a^*)$, that is, the comparison of an observed contingency table with an estimated contingency table, are called measures of interaction or goodness-of-fit. Measures of the form $2I(x_b^*:x_a^*)$, comparing two estimated contingency tables, are called measures of effect, that is the effect of the marginals in the set H_b but not in the set H_a or the taus in x_b^* but not in x_a^*. We note that $2I(x:x_a^*)$ tests a null hypothesis that the values of the τ parameters in the representation of the observed contingency table $x(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero and the number of these taus is the number of degrees of freedom. Similarly $2I(x_b^*:x_a^*)$ tests a null hypothesis that the values of the τ parameters in the representation of the estimated table $x_a^*(\omega)$ are zero and the number of these taus is the number of degrees of freedom.

We summarize the additive relationships of the m.d.i. statistics and the associated degrees of freedom in the Analysis of Information Table 9.1.

TABLE 9.1

ANALYSIS OF INFORMATION TABLE

Component due to	Information	D.F.
H_a : Interaction	$2I(x:x_a^*)$	N_a
H_b : Effect	$2I(x_b^*:x_a^*)$	$N_a - N_b$
Interaction	$2I(x:x_b^*)$	N_b

Since measures of the form $2I(x:x_a^*)$ may also be interpreted as measures of the "variation unexplained" by the estimate x_a^*, the additive relationship leads to the interpretation of the ratio

$$(9.2) \qquad \frac{2I(x:x_a^*) - 2I(x:x_b^*)}{2I(x:x_a^*)} = \frac{2I(x_b^*:x_a^*)}{2I(x:x_a^*)}$$

as the percentage of the unexplained variation due to x_a^* accounted for by the additional constraints defining x_b^*. The ratio (9.2) is thus similar to the squared correlation coefficients associated with normal distributions.

We remark that the marginals explicit and implicit of the estimated table $x_a^*(\omega)$ which form the set of restraints H_a used to generate $x_a^*(\omega)$ are the same as the corresponding marginals of the observed $x(\omega)$ table and all lower order implied marginals. It may be shown that $2I(x:x_a^*)$ is approximately a quadratic in the differences between the remaining marginals of the $x(\omega)$ table and the corresponding ones as calculated from the $x_a^*(\omega)$ table.

Similarly $2I(x_b^*:x_a^*)$ is also approximately a quadratic in the differences between those additional marginal restraints in H_b but not in H_a and the corresponding marginal values as computed from the $x_a^*(\omega)$ table.

As may be seen, because of the nature of the $T(\omega)$ functions described in Section 7.1 or indicated in Figure 8.1, the τ's are determined from the log-linear regression Equations (6.7) (see (8.2) and (11.3))

as sums and differences of values of $\ell n\ x^*(ijk\ell)$. A variety of statistics have been presented in the literature for the analysis of contingency tables which are quadratics in differences of marginal values or quadratics in the τ's or the linear combinations of logarithms of the observed or estimated values. The principle of minimum discrimination information estimation and its procedures thus provides a unifying relationship since such statistics may be seen as quadratic approximations of the minimum discrimination information statistic. We remark that the corresponding approximate X^2's are not generally additive.

We mention the approximations in terms of quadratic forms in the marginals or the τ's as a possible bridge connecting the familiar procedures of classical regression analysis and the procedures proposed here to assist in understanding and interpreting the analysis of information tables. The covariance matrix of the $T(\omega)$ functions or the taus can be obtained for either the observed table or any of the estimated tables, as well as the inverse matrices as part of the output of the general computer program.

10. Outliers

We define outliers as observations in one or more cells of a contingency table which apparently deviate significantly from a fitted model. These outliers may lead one to reject a model which fits the other observations. For example, in multi-dimensional contingency tables in which time or age is one of the classifications there may occur an age effect such that a model may be rejected for the entire table but a model taking the possible age effect into account may lead to an acceptable partitioning of the model.

In other cases even though a model seems to fit, the outliers contribute much more than reasonable to the measure of deviation between the data and the fitted values of the model. In other words, the outliers make up a large percentage of the "unexplained variation" $2I(x:x^*)$.

A clue to possible outliers is provided by the output of the computer program. In the computer output for each estimate five entries are

listed for each cell. The fourth of these is titled OUTLIER and its numerical value provides a lower bound for the decrease in the corresponding $2I(x:x^*)$ if that cell were not included in the fitting procedure. Since the reduction in the degrees of freedom is one for each omitted cell, values of OUTLIER greater than say 3.5 are of interest. The basis for the OUTLIER computation and interpretation follows. Let x_a^* denote the minimum discrimination information estimate subject to certain marginal restraints. Let x_b^* denote the minimum discrimination information estimate subject to the same marginal restraints as x_a^* except that the value $x(\omega_1)$, say, is not included, so that $x_b^*(\omega_1) = x(\omega_1)$. The basic additivity property of the minimum discrimination information statistics states that

$$2I(x:x_a^*) = 2I(x_b^*:x_a^*) + 2I(x:x_b^*)$$

or

$$2I(x:x_a^*) - 2I(x:x_b^*) = 2I(x_b^*:x_a^*) .$$

These results are summarized in the Analysis of Information Table 10.1.

TABLE 10.1

ANALYSIS OF INFORMATION TABLE

Component due to	Information	D.F.
H_a :	$2I(x:x_a^*)$	N_a
H_b : Same as H_a but omitting $x(\omega_1)$	$2I(x_b^*:x_a^*)$	1
	$2I(x:x_b^*)$	$N_b = N_a - 1$

But

$$2I(x_b^*:x_a^*) = 2\left(x_b^*(\omega_1)\ \ln \frac{x_b^*(\omega_1)}{x_a^*(\omega_1)} + \sum_{\Omega-\omega_1} x_b^*(\omega)\ \ln \frac{x_b^*(\omega)}{x_a^*(\omega)}\right)$$

(10.1)

$$= 2\left(x(\omega_1)\ \ln \frac{x(\omega_1)}{x_a^*(\omega_1)} + \sum_{\Omega-\omega_1} x_b^*(\omega)\ \ln \frac{x_b^*(\omega)}{x_a^*(\omega)}\right) ,$$

and using the convexity property which implies that

$$(10.2)\qquad \sum_{\Omega-\omega_1} x_b^*(\omega)\ \ln \frac{x_b^*(\omega)}{x_a^*(\omega)} \geq \left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right) \ln \frac{\left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right)}{\left(\sum_{\Omega-\omega_1} x_a^*(\omega)\right)}$$

$$= (n - x_b^*(\omega_1))\ \ln \frac{n - x_b^*(\omega_1)}{n - x_a^*(\omega_1)} ,$$

we get from (10.1) that

$$(10.3)\qquad 2I(x_b^*:x_a^*) \geq 2\left(x(\omega_1)\ \ln \frac{x(\omega_1)}{x_a^*(\omega_1)} + \left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right) \ln \frac{\left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right)}{\left(\sum_{\Omega-\omega_1} x_a^*(\omega)\right)}\right)$$

$$= 2\left(x(\omega_1)\ \ln \frac{x(\omega_1)}{x_a^*(\omega_1)} + (n - x(\omega_1))\ln \frac{n - x(\omega_1)}{n - x_a^*(\omega_1)}\right).$$

The last value can be computed and is listed as the OUTLIER entry for each cell of the computer output for the estimate x_a^* .

The ratio

$$(10.4)\qquad \frac{2I(x:x_a^*) - 2I(x:x_b^*)}{2I(x:x_a^*)} = \frac{2I(x_b^*:x_a^*)}{2I(x:x_a^*)}$$

then indicates the percentage of the "unexplained variation" due to the outlier value.

11. The 2x2 Table

It may be useful to reexamine the 2x2 table from the point of view of the preceding discussion. The algebraic details are simple in this case and exhibit the unification of the information theoretic development.

Suppose we have the observed 2x2 table in Figure 11.1.

x(11)	x(12)	x(1.)
x(21)	x(22)	x(2.)
x(.1)	x(.2)	n

Figure 11.1.

If we obtain the m.d.i. estimate fitting the one-way marginals, the generalized independence hypothesis is the classical independence hypothesis and the minimum discrimination information estimate is $x^*(ij) = x(i.)x(.j)/n$. The representation of the log-linear regression (6.7) as in Figure 8.1 for the full model is given in Figure 11.2. The entries in the columns τ_1 , τ_2 , τ_3

i	j	L	τ_1	τ_2	τ_3
1	1	1	1	1	1
1	2	1	1		
2	1	1		1	
2	2	1			

Figure 11.2.

are, respectively, the values of the functions $T_1(ij)$, $T_2(ij)$, $T_3(ij)$ associated with the marginals $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$, and the column headed L corresponds to the normalizing factor (the negative of the logarithm of the moment-generating function as in (6.7)).

We recall the interpretation of Figure 11.2 as the log-linear relations

$$(11.1)\quad \begin{cases} \ln \dfrac{x(11)}{n\pi} = L + \tau_1 + \tau_2 + \tau_3 \\ \ln \dfrac{x(12)}{n\pi} = L + \tau_1 \\ \ln \dfrac{x(21)}{n\pi} = L + \tau_2 \\ \ln \dfrac{x(22)}{n\pi} = L \; . \end{cases}$$

From (11.1) we find

$$(11.2)\quad \begin{aligned} L &= \ln\, (x(22)/n/4)\ , \\ \tau_1 &= \ln\, (x(12)/x(22))\ , \\ \tau_2 &= \ln\, (x(21)/x(22))\ , \\ \tau_3 &= \ln\, (x(11)x(22)/x(12)x(21)) \end{aligned}$$

or

$$\begin{aligned}\tau_1 &= \ln x(12) - \ln x(22) \ ,\\ \tau_2 &= \ln x(21) - \ln x(22) \ ,\\ \tau_3 &= \ln x(11) + \ln x(22) - \ln x(12) - \ln x(21) \ .\end{aligned} \tag{11.3}$$

If we call $\underline{T}$ the matrix with columns the columns of the design matrix of Figure 11.2, that is,

$$\underline{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} , \tag{11.4}$$

and define a diagonal matrix $\underline{D}$ with main diagonal the elements $x(ij)$, that is,

$$\underline{D} = \begin{pmatrix} x(11) & 0 & 0 & 0 \\ 0 & x(12) & 0 & 0 \\ 0 & 0 & x(21) & 0 \\ 0 & 0 & 0 & x(22) \end{pmatrix} , \tag{11.5}$$

then the estimate of the covariance matrix of $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$ for the observed contingency table is $\underline{\Sigma} = \underline{A}_{22.1}$ where

$$\underline{A} = \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} = \underline{T}'\underline{DT} \tag{11.6}$$

$$\underline{A}_{22.1} = \underline{A}_{22} - \underline{A}_{21}\,\underline{A}_{11}^{-1}\,\underline{A}_{12} \tag{11.7}$$

and $\underline{A}_{11}$ is 1×1, $\underline{A}_{22}$ is 3×3, $\underline{A}_{21}' = \underline{A}_{12}$ is 1×3. It is found that

$$\underline{\Sigma} = \begin{pmatrix} \frac{x(1.)x(2.)}{n} & x(11) - \frac{x(1.)x(.1)}{n} & \frac{x(11)x(2.)}{n} \\ x(11) - \frac{x(1.)x(.1)}{n} & \frac{x(.1)x(.2)}{n} & \frac{x(11)x(.2)}{n} \\ \frac{x(11)x(2.)}{n} & \frac{x(11)x(.2)}{n} & x(11) - \frac{x^2(11)}{n} \end{pmatrix} , \tag{11.8}$$

and the inverse matrix is

$$(11.9)\quad \underline{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{x(12)} + \frac{1}{x(22)} & \frac{1}{x(22)} & -\frac{1}{x(12)} - \frac{1}{x(22)} \\ \frac{1}{x(22)} & \frac{1}{x(21)} + \frac{1}{x(22)} & -\frac{1}{x(21)} - \frac{1}{x(22)} \\ -\frac{1}{x(12)} - \frac{1}{x(22)} & -\frac{1}{x(21)} - \frac{1}{x(22)} & \frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)} \end{pmatrix} .$$

We remark that the matrix in (11.9) is the covariance matrix of the τ's in (11.3).

Note that the value of the logarithm of the cross-product ratio, a measure of association or interaction, appears in the course of the analysis as the value of τ_3 for the observed values $x(ij)$, and that $\tau_3 = 0$ for $x^*(ij)$, the estimate under the hypothesis of independence, for which the representation as in Figure 11.2 does not involve the last column since it is obtained by fitting the one-way marginals.

The log-linear relations for the estimate $x^*(ij)$ are

$$(11.10)\quad \begin{cases} \ln \dfrac{x^*(11)}{n\pi} = L + \tau_1 + \tau_2 \\ \ln \dfrac{x^*(12)}{n\pi} = L + \tau_1 \\ \ln \dfrac{x^*(21)}{n\pi} = L + \tau_2 \\ \ln \dfrac{x^*(22)}{n\pi} = L \end{cases} ,$$

where the numerical values of L, τ_1, τ_2 in (11.10) depend on x^* and differ from the values in (11.1).

The minimum discrimination information statistic to test the null hypothesis or model of independence is $2I(x:x^*)$ with one degree of freedom. In this case the quadratic approximation is

$$(11.11)\quad 2I(x:x^*) \approx \left(x(11) - \frac{x(1.)x(.1)}{n}\right)^2 \left(\frac{1}{x^*(11)} + \frac{1}{x^*(12)} + \frac{1}{x^*(21)} + \frac{1}{x^*(22)}\right) .$$

Remembering that $x^*(ij) = x(i.)x(.j)/n$, the right-hand side of (11.11) may also be shown to be

$$(11.12) \qquad \chi^2 = \Sigma\ (x(ij) - x(i.)x(.j)/n)^2 \Big/ \frac{x(i.)x(.j)}{n} ,$$

the classical χ^2-test for independence with one degree of freedom. Another test which has been proposed for the null hypothesis of no association or no interaction in the 2x2 table is

$$(11.13) \quad (\ln x(11) + \ln x(22) - \ln x(12) - \ln x(21))^2 \left(\frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)}\right)^{-1} ,$$

which may be shown to be a quadratic approximation for $2I(x:x^*)$ in terms of τ_3 with the covariance matrix estimated using the observed values and not the estimated values. We remark that if the observed values are used to estimate the covariance matrix then instead of the classical χ^2-test in (11.12) there is derived the modified Neyman chi-square

$$(11.14) \qquad \chi_1^2 = \Sigma\ (x(ij) - x(i.)x(.j)/n)^2/x(ij) .$$

12. An Analysis

In order to coordinate and relate the various definitions, concepts, parameters, computational features, etc. discussed in the preceding sections we shall consider in detail the analysis of a specific contingency table.

Table 12.1 is a four-way contingency table of 14,053 marines who enlisted in 1966 or 1967, cross-classified on the variables home of record, level of education, race and boot camp completion. We denote the occurrences in the four-way cross-classification or contingency Table 12.1 by $x(ijk\ell)$ with the notation

Variable	Index	1	2	3	4
Home of Record	i	East	North	West	South
Level of Education	j	Below H.S.	H.S.	Above H.S.	
Race	k	White	Non-white		
Boot Camp	ℓ	Failed	Passed		

TABLE 12.1

BOOT CAMP COMPLETION

$x(ijk\ell)$

	i	East						North					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	62	10	44	4	8	0	20	4	18	2	6	0
	P	944	133	1881	195	320	12	870	103	2205	148	471	24

	i	West						South					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	14	2	9	0	0	1	42	9	34	16	6	0
	P	567	40	1350	94	421	19	952	228	1741	508	461	55

For this data we are interested in the possible relationship of success in boot camp as a dependent variable on the independent or explanatory variables home of record, level of education, and race. To obtain a smoothed estimate of the observed cross-classification utilizing significant effects and interactions we shall examine a sequence of minimum discrimination information estimates based on nested sets of fitted marginals. That is, each successive estimate uses a set of marginals which explicitly or implicitly contains the marginals of the preceding estimate and also additional ones to determine the effect of the additional marginals or their associated interaction tau parameters. The analysis of information table permits us to judge the significance or non-significance of these effects or interaction tau parameters.

12.1. <u>Fitting Nested Sets of Marginals</u>. Since we are interested in the possible relationship of success in boot camp on home of record, level of education and race, we first fit the marginals $x(ijk.)$, $x(\ldots\ell)$ since the corresponding estimate $x^*(ijk\ell) = x(ijk.)x(\ldots\ell)/n$ is that under the null hypothesis or model of independence of success and the joint variable (home of record, level of education, race) or no interaction between success and the joint variable. In other words we first want to determine whether the 24 columns of Table 12.1 are homogeneous or not with respect to the underlying probabilities of passing or failing. The associated m.d.i. statistic is

$$2I(x:x^*) = 2\ \Sigma\ \Sigma\ \Sigma\ \Sigma\ x(ijk\ell)\ \ell n\big(x(ijk\ell)/x^*(ijk\ell)\big) = 160.551$$

with 23 degrees of freedom. We reject the hypothesis of independence or no interaction. We therefore shall look for explanatory effects.

In Figure 12.1 there is given the complete schematic for the log-linear representations. The representation for the estimate of joint independence $x^*(ijk\ell) = x(ijk.)x(\ldots\ell)/n$ uses columns 1-17, 21-22, 26-31 corresponding to all the marginals explicit and implicit in the fitted marginal constraints. We can also interpret $2I(x:x^*)$ as testing a null hypothesis or model that the 23 tau parameters in the representation of x but not in x^* are zero, that is, the parameters corresponding to columns [illegible]-20, 23-25, 32-48.

i j k l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
	L	i 1	i 2	i 3	j 1	j 2	k 1	l 1	ij 11	ij 12	ij 21	ij 22	ij 31	ij 32	ik 11	ik 21	ik 31	il 11	il 21	il 31	jk 11	jk 21	jl 11	jl 21	kl 11	ijk 111	ijk 121	ijk 211	ijk 221	ijk 311	ijk 321	ijl 111	ijl 121	ijl 211	ijl 221	ijl 311	ijl 321	ikl 111	ikl 211	ikl 311	jkl 111	jkl 211	ijkl 1111	ijkl 1211	ijkl 2111	ijkl 2211	ijkl 3111	ijkl 3211
1 1 1 1	1	1			1		1	1	1					1			1			1		1		1	1							1						1			1		1					
1 1 1 2	1	1			1		1		1						1							1					1																						
1 1 2 1	1	1			1			1	1									1					1										1																
1 1 2 2	1	1			1				1																																							
1 2 1 1	1	1				1	1	1		1					1			1				1		1	1		1						1					1				1		1				
1 2 1 2	1	1				1	1			1					1							1					1																					
1 2 2 1	1	1				1		1		1								1						1									1															
1 2 2 2	1	1				1				1																																						
1 3 1 1	1	1					1	1							1			1							1													1										
1 3 1 2	1	1					1								1																																
1 3 2 1	1	1						1										1																													
1 3 2 2	1	1																																													
2 1 1 1	1		1		1		1	1			1					1			1		1		1		1			1						1					1		1				1			
2 1 1 2	1		1		1		1				1					1					1							1																			
2 1 2 1	1		1		1			1			1								1				1											1														
2 1 2 2	1		1		1						1																																				
2 2 1 1	1		1			1	1	1				1				1			1			1		1	1				1						1				1			1				1		
2 2 1 2	1		1			1	1					1				1						1							1																		
2 2 2 1	1		1			1		1				1							1					1											1													
2 2 2 2	1		1			1						1																																			
2 3 1 1	1		1				1	1							1			1						1														1									
2 3 1 2	1		1				1								1																																
2 3 2 1	1		1					1										1																													
2 3 2 2	1		1																																												
3 1 1 1	1			1	1		1	1					1				1			1	1		1		1					1						1				1	1						1	
3 1 1 2	1			1	1		1						1				1				1									1																		
3 1 2 1	1			1	1			1					1							1			1													1												
3 1 2 2	1			1	1								1																																		
3 2 1 1	1			1		1	1	1						1			1			1		1		1	1						1						1			1		1						1
3 2 1 2	1			1		1	1							1			1					1									1																	
3 2 2 1	1			1		1		1						1						1				1													1											
3 2 2 2	1			1		1								1																																	
3 3 1 1	1			1			1	1									1			1					1															1								
3 3 1 2	1			1			1										1																														
3 3 2 1	1			1				1												1																												
3 3 2 2	1			1																																												
4 1 1 1	1				1		1	1													1		1		1																1							
4 1 1 2	1				1		1														1																											
4 1 2 1	1				1			1															1																									
4 1 2 2	1				1																																											
4 2 1 1	1					1	1	1														1		1	1																	1						
4 2 1 2	1					1	1															1																										
4 2 2 1	1					1		1																1																								
4 2 2 2	1					1																																										
4 3 1 1	1						1	1																	1																							
4 3 1 2	1						1																																									
4 3 2 1	1							1																																								
4 3 2 2	1																																															

Figure 12... Log-linear Representation

The value of $2I(x:x^*)$ is so large that we reject the model of joint independence. We therefore proceed to fit a sequence of nested marginals all including $x(ijk.)$ and various combinations of two- and three-way marginals containing success with other variables. We summarize some results in the truncated Analysis of Information Table 12.2. We have not included all the intermediate fitting sequences for conciseness. We remark that although the measure of the effect of additional marginals or their associated parameters may vary according to the sequence in which they have been added, significant effects tend to remain significant and non-significant effects tend to stay non-significant so that the first overall survey should determine the estimates and interaction parameters which warrant further investigation. For example, the effect of adding $x(..k\ell)$ to $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$ is given in Analysis of Information Table 12.3 as $2I(x_f^*:x_a^*) = 1.410$ with one degree of freedom, but the effect of adding $x(..k\ell)$ to $x(ijk.)$, $x(ij.\ell)$ is given in Analysis of Information Table 12.2 as $2I(x_e^*:x_m^*) = 1.239$ with one degree of freedom. In neither case is the effect or the corresponding tau parameter $\tau_{11}^{k\ell}$ significant.

The columns of Figure 12.1 which occur in the log-linear representations of the estimates retained in Analysis of Information Table 12.2 are

Marginals Fitted	Estimate	Columns of Figure 12.1
$x(ijk.)$, $x(...\ell)$	x^*	1-17, 21-22, 26-31
$x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$	x_a^*	1-24, 26-31
$x(ijk.)$, $x(ij.\ell)$	x_m^*	1-24, 26-37
$x(ijk.)$, $x(ij.\ell)$, $x(..k\ell)$	x_e^*	1-37 .

From the analytic form of the log-linear representation or by taking differences of appropriate rows of Figure 12.1 within the columns used for the estimate, the log-odds of fail to pass for each of the estimates are given by the respective parametric representations in (12.1) where the superscripts relate to the variables and the subscripts range over the possible indices. The values of the parameters depend of course on the corresponding estimate.

TABLE 12.2

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
$x(ijk.), x(...\ell)$	$2I(x:x^*) = 160.551$	23
a) $x(ijk.), x(i..\ell), x(.j.\ell)$	$2I(x_a^*:x^*) = 138.732$	5
	$2I(x:x_a^*) = 21.819$	18
m) $x(ijk.), x(ij.\ell)$	$2I(x_m^*:x_a^*) = 7.384$	6
	$2I(x:x_m^*) = 14.435$	12
e) $x(ijk.), x(ij.\ell), x(..k\ell)$	$2I(x_e^*:x_m^*) = 1.239$	1
	$2I(x:x_e^*) = 13.196$	11

$$\frac{2I(x:x^*) - 2I(x:x_a^*)}{2I(x:x^*)} = \frac{138.732}{160.551} = 0.86$$

$$\frac{2I(x:x^*) - 2I(x:x_m^*)}{2I(x:x^*)} = \frac{146.116}{160.551} = 0.91$$

$$\frac{2I(x:x^*) - 2I(x:x_e^*)}{2I(x:x^*)} = \frac{147.355}{160.551} = 0.92$$

TABLE 12.3

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
a) $x(ijk.), x(i..\ell), x(.j.\ell)$	$2I(x:x_a^*) = 21.819$	18
f) $x(ijk.), x(i..\ell), x(.j.\ell), x(..k\ell)$	$2I(x_f^*:x_a^*) = 1.410$	1
	$2I(x:x_f^*) = 20.409$	17

$$\ell n \frac{x_a^*(ijk1)}{x_a^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell}$$

$$\ell n \frac{x_m^*(ijk1)}{x_m^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{ij1}^{ij\ell} \tag{12.1}$$

$$\ell n \frac{x_e^*(ijk1)}{x_e^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{11}^{k\ell} + \tau_{ij1}^{ij\ell}$$

We recall that parameters with indices $i = 4$ and/or $j = 3$ and/or $k = 2$ and/or $\ell = 2$ are by convention set equal to zero.

We remark that $x_m^*(ijk\ell)$, determined by fitting the marginals $x(ijk.)$, $x(ij.\ell)$, is expressible explicitly as

$$x_m^*(ijk\ell) \doteq x(ijk.)x(ij.\ell)/x(ij..) \tag{12.2}$$

and is the estimate under a null hypothesis that race and success are conditionally independent given home of record and level of education. In Analysis of Information Table 12.2 the value $2I(x:x_m^*) = 14.435$, 12 degrees of freedom, indicates an acceptable fit of this model. Furthermore, $2I(x_e^*:x_m^*) = 1.239$, one degree of freedom, implies that the additional effect of the marginal $x(..k\ell)$ is not significant or that in the parametric representation of the log-odds in (12.1) the parameter $\tau_{11}^{k\ell}$ measuring the effect of race on the dependent variable success is not significant. We therefore investigate the estimate x_m^* in greater detail. The values of $x_m^*(ijk\ell)$ are given in Table 12.4.

In the expression for the log-odds under x_m^* in (12.1) τ_1^{ℓ} is an overall average, $\tau_{i1}^{i\ell}$ and $\tau_{j1}^{j\ell}$ are the effects of home of record and level of education on boot camp completion and $\tau_{ij1}^{ij\ell}$ is the interaction effect of home of record x level of education on boot camp completion. The numerical values of the tau parameters are given in Table 12.5. We recall that by convention parameters with an index corresponding to $i = 4$ and/or $j = 3$ and/or $\ell = 2$ are equal to zero.

TABLE 12.4

BOOT CAMP COMPLETION

$x_m^*(ijk\ell)$

i		East						North					
j		Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
k		W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	63.039	8.961	43.503	4.497	7.718	0.282	21.424	2.576	18.736	1.264	5.713	0.287
	P	942.960	134.039	1881.497	194.503	320.282	11.718	868.575	104.424	2204.264	148.736	471.287	23.713

i		West						South					
j		Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
k		W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	14.921	1.079	8.418	0.582	0.955	0.045	41.181	9.819	38.604	11.396	5.368	0.632
	P	566.078	40.921	1350.582	93.418	420.045	19.955	952.819	227.181	1736.396	512.604	461.632	54.368

TABLE 12.5

VALUES OF PARAMETERS IN LOG-ODDS FOR x_m^* IN (11.1)

$\tau_1^{\ell} = -4.454347$	$\tau_{111}^{ij\ell} = -0.292478$
$\tau_{11}^{i\ell} = 0.728653$	$\tau_{121}^{ij\ell} = -0.689433$
$\tau_{21}^{i\ell} = 0.041549$	$\tau_{211}^{ij\ell} = -0.602435$
$\tau_{31}^{i\ell} = -1.632427$	$\tau_{221}^{ij\ell} = -1.003045$
$\tau_{11}^{j\ell} = 1.312903$	$\tau_{311}^{ij\ell} = 1.137932$
$\tau_{21}^{j\ell} = 0.648130$	$\tau_{321}^{ij\ell} = 0.360697$

From the parametric representation of the log-odds in (12.1) and the values in Table 12.5 one can determine differences in the log-odds associated with changes in various categories. Thus the differences in the log-odds (fail to pass) as one changes the home of record, for fixed level of education, are given by

	E-N	E-W	E-S
Below H.S.	0.9970	0.7287	0.4362
H.S.	1.0007	1.3110	0.0392
Above H.S.	0.6871	2.3611	0.7287

The differences in the log-odds as one changes the level of education for fixed home of record are given by

	Below H.S.-H.S.	H.S.-Above H.S.
East	1.0617	-0.0413
North	1.0654	-0.3549
West	1.4420	1.0088
South	0.6648	0.6481

For easier interpretation, we convert the log-odds values to ratios of the odds of failure.

	E-N	E-W	E-S
Below H.S.	2.7	2.1	1.6
H.S.	2.7	3.7	1.0
Above H.S.	2.0	10.6	2.1

	Below H.S.-H.S.	H.S.-Above H.S.
East	2.9	0.96
North	2.9	0.70
West	4.2	2.7
South	1.9	1.9

Note that the odds of failure in boot camp of a recruit with home of record East and Above H.S. level of education are 10.6 times the odds of a recruit with the same level of education but home of record West. Recruits with home of record East or North but with level of education H.S. do better than recruits with same home of record but Above H.S. level of education.

We have also computed the odds of failure $x^*_m(ijk1)/x^*_m(ijk2)$ and listed the results in increasing values. The odds are expressed to 1,000, that is, 5 to 1,000, 6 to 1,000, etc.

Home of Record	Level of Education	Odds
West	Above H.S.	2
West	H.S.	6
North	H.S.	9
South	Above H.S.	12
North	Above H.S.	12
South	H.S.	22
East	H.S.	23
East	Above H.S.	24
North	Below H.S.	25
West	Below H.S.	26
South	Below H.S.	43
East	Below H.S.	67

Note that the overall odds of failure for this data are 311/13742 = 0.0226 or 23.

For ease of comparison and inference, we also list the foregoing results by home of record and level of education.

	West	North	South	East
Above H.S.	2	12	12	24
H.S.	6	9	22	23
Below H.S.	26	25	43	67

Examination of the computer output for $x_m^*(ijk\ell)$ shows that for West, Above H.S., Non-white, Fail, the value of OUTLIER is 4.28. From Table 12.1, we see that the corresponding observed values are given by the two-way table

West, Above H.S., $x(33k\ell)$

	White	Non-white	
Fail	0	1	1
Pass	421	19	440
	421	20	441

and from Table 12.4,the corresponding estimated values are

West, Above H.S., $x_m^*(33k\ell)$

	White	Non-white	
Fail	0.955	0.045	1.000
Pass	420.045	19.955	440.000
	421.000	20.000	441.000

Testing the observed two-way table West, Above H.S., $x(33k\ell)$ for independence of race and boot camp completion by the statistic

$$2 \sum_k \sum_\ell x(33k\ell)\ \ell n\left(x(33k\ell)\Big/\frac{x(33k.)\,x(33.\ell)}{x(33..)}\right) = 2\left\{\sum_k \sum_\ell x(33k\ell)\ \ell n\ x(33k\ell)\right.$$

$$\left. + x(33..)\ \ell n\ x(33..) - \sum_k x(33k.)\ \ell n\ x(33k.) - \sum_\ell x(33.\ell)\ \ell n\ x(33.\ell)\right\}$$

yields the value 6.236, one degree of freedom. (Tables of $2n \ln n$, n an integer 1 to 10,000, are available for such calculations.) The contribution of West, Above H.S. to the value of $2I(x:x_m^*)$ is obtained by the computer as

$$2\left(0 \ln \frac{0}{0.955} + 1 \ln \frac{1}{0.045} + 421 \ln \frac{421}{420.045} + 19 \ln \frac{19}{19.955}\right)$$

and yields the same value 6.236.

Because the value 6.236 is statistically significant at the 0.02 level, the OUTLIER statistic has shown an "unusual" situation for $x_m^*(ijk\ell)$ corresponding to West, Above H.S.

We shall consider the procedure to account for outliers after we examine the estimate x_a^*.

In view of the fact that the Analysis of Information Table 12.2 shows no significant effects for the estimates following x_a^* and since $2I(x:x_a^*) = 21.815$, 18 degrees of freedom, implies an acceptable fit, let us examine the estimate x_a^* with possible outliers in mind. The values of the estimate x_a^* are given in Table 12.6.

The log-odds of fail to pass for x_a^* are given in (12.1) with the parameters having the same interpretation as those for x_m^* except that there is no interaction effect. The values of the parameters for x_a^* are given in Table 12.7.

For the estimate x_a^* the ratio of the odds of failure between different homes of record is the same for all levels of education and, of course, the ratio of the odds of failure for different educational levels is the same for all homes of records For the ratio of odds, and odds, see Tables 12.8, 12.9 and 12.10.

Examination of the computer output for x_a^* shows an OUTLIER value of 5.20 for West, Above H.S., White, Fail and an OUTLIER value of 3.54 for South, H.S., Non-white, Fail. The corresponding observed and estimated cell entries are

TABLE 12.6

$x_a^*(ijk\ell)$

	i	East						North					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	61.099	8.685	46.688	4.826	6.464	0.237	21.381	2.571	20.837	1.406	3.623	0.182
	P	944.901	134.315	1878.312	194.174	321.535	11.763	868.619	104.429	2202.163	148.594	473.377	23.818

	i	West						South					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	11.381	0.823	10.358	0.716	2.599	0.123	46.075	10.986	32.557	9.611	6.953	0.819
	P	569.619	41.177	1348.642	93.284	418.401	19.877	947.925	226.014	1742.443	514.389	460.047	54.181

TABLE 12.7

PARAMETER VALUES IN LOG-ODDS REPRESENTATION

	x_a^*	x_b^*	x_c^*
τ_1^{ℓ}	-4.192224	-4.059831	-4.105023
$\tau_{11}^{i\ell}$	0.285423	0.288534	0.364671
$\tau_{21}^{i\ell}$	-0.680394	-0.680769	-0.602516
$\tau_{31}^{i\ell}$	-0.889058	-0.771589	-0.690762
$\tau_{11}^{j\ell}$	1.168221	1.025047	1.019191
$\tau_{21}^{j\ell}$	0.212164	0.067678	-0.001819

TABLE 12.8

RATIOS OF THE ODDS OF FAILURE

	x_a^*	x_b^*	x_c^*
East/South	1.33	1.33	1.44 (H.S.; Non-white 0.75)
North/South	0.51	0.51	0.55 (H.S.; Non-white 0.29)
West/South	0.41	0.46	0.50 (H.S.; Non-white 0.26)
Below H.S./H.S.	2.60	2.61	2.78 (South, White) 1.45 (South, Non-white)
H.S./Above H.S.	1.24	1.07 (West, Non-white)	1.00 (West, Non-white, South, White) 1.91 (South, Non-white)

TABLE 12.9

ODDS OF FAILURE, EXPRESSED TO 1,000

Home of Record	Level of Education	Odds		
		$x_a^*(ijk1)/x_a^*(ijk2)$	$x_b^*(ijk1)/x_b^*(ijk2)$	$x_c^*(ijk1)/x_c^*(ijk2)$
West	Above H.S.	6	0 White, 8 Non-white	0 White, 8 Non-white
North	Above H.S.	8	9	9
West	H.S.	8	9	8
North	H.S.	9	9	9
South	Above H.S.	15	17	16
South	H.S.	19	18	16 White, 32 Non-white
West	Below H.S.	20	22	23
East	Above H.S.	20	23	24
North	Below H.S.	25	24	25
East	H.S.	25	25	24
South	Below H.S.	49	48	46
East	Below H.S.	65	64	66

TABLE 12.10

ODDS OF FAILURE, EXPRESSED TO 1,000

		West	North	South	East
Above H.S.	x_a^*	6	8	15	20
	x_b^*	0 White, 8 Non-white	9	17	23
	x_c^*	0 White, 8 Non-white	9	16	24
H.S.	x_a^*	8	9	19	25
	x_b^*	9	9	18	25
	x_c^*	8	9	16 White, 32 Non-white	24
Below H.S.	x_a^*	20	25	49	65
	x_b^*	22	24	48	64
	x_c^*	23	25	46	66

West, Above H.S.

$x(33k\ell)$

	White	Non-white
Fail	0	1
Pass	421	19
	421	20

$x_a^*(33k\ell)$

	White	Non-white
Fail	2.599	0.123
Pass	418.401	19.877
	421.000	20.000

South, H.S.

$x(42k\ell)$

	White	Non-white
Fail	34	16
Pass	1741	508
	1775	524

$x_a^*(42k\ell)$

	White	Non-white
Fail	32.557	9.611
Pass	1742.443	514.389
	1775.000	524.000

12.2. The Estimate $x_m^*(ijk\ell)$ Adjusted for Outliers. For all estimates considered under the nested marginal hypotheses, a requirement was that $x^*(ijk.) = x(ijk.)$. Accordingly for the model with interaction we require the modified estimate to be fitted using the marginals $x(ijk.)$, $x(ij.\ell)$ derived from all observations except the outliers $x(3311)$ and $x(3312)$. We shall use the observed values as the estimates for the outlier cells. Thus if we denote the modified estimates by $x_r^*(ijk\ell)$ we have $x_r^*(3311) = x(3311)$ and $x_r^*(3312) = x(3312)$.

Because of the marginals used for fitting, it turns out that the values of the modified estimate, $x_r^*(ijk\ell)$ are equal to the values of the original estimate $x_m^*(ijk\ell)$ (since $x_r^*(ijk\ell) = x(ijk.)x(ij.\ell)/x(ij..)$) except, of course, for the cells (3311) and (3312) , and to satisfy the requirement that $x_r^*(ij.\ell) = x(ij.\ell)$ it follows that $x_r^*(3321) = x(3321)$, $x_r^*(3322) = x(3322)$. The associated Analysis of Information Table 12.11 follows.

TABLE 12.11

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
m) $x(ijk.)$, $x(ij.\ell)$	$2I(x:x_m^*) = 14.435$	12
r) $x(ijk.)$, $x(ij.\ell)$ less $x(3311)$, $x(3312)$	$2I(x_r^*:x_m^*) = 6.235$	1
	$2I(x:x_r^*) = 8.200$	11

Note that since $x_r^*(ijk\ell) = x_m^*(ijk\ell)$ except that $x_r^*(3311) = x(3311)$, $x_r^*(3312) = x(3312)$, $x_r^*(3321) = x(3321)$, $x_r^*(3322) = x(3322)$, the value of the measure of effect $2I(x_r^*:x_m^*)$ is the same as that earlier derived in the test for conditional independence.

The global inference that race and boot camp completion are conditionally independent is valid except for West, Above H.S., and with the estimate x_r^* the odds of failure for White are zero whereas they are 53 in 1,000 for Non-white.

Since $2I(x_r^*:x_m^*)/2I(x:x_m^*) = 6.235/14.435 = 0.43$, we conclude that the outlier value West, Above H.S. accounts for 43% of the "unexplained variation" $2I(x:x_m^*)$.

12.3. The Estimate $x_a^*(ijk\ell)$ Adjusted for Outliers. We shall first derive a revised estimate for $x_a^*(ijk\ell)$ adjusted for the outlier $x(3311)$, $x(3312)$, that is, we fit the marginals $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$ excluding the observations $x(3311)$, $x(3312)$ (West, Above H.S., White, Fail; West, Above H.S., White, Pass). Thus if we denote the new estimate by $x_b^*(ijk\ell)$ we have $x_b^*(3311) = x(3311)$, $x_b^*(3312) = x(3312)$. The values of the estimate x_b^* are given in Table 12.12.

In particular, note that for West, Above H.S., the corresponding observed and estimated cell entries are

TABLE 12.12

$x_b^*(ijk\ell)$

	i	East						North					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	60.661	8.623	46.281	4.784	7.381	0.270	21.150	2.543	20.581	1.389	4.130	0.208
	P	945.339	134.377	1878.718	194.216	320.619	11.730	868.850	104.457	2202.418	148.611	472.870	23.792

	i	West						South					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	12.634	0.913	11.499	0.795	0	0.158	45.604	10.873	32.172	9.498	7.920	0.933
	P	568.366	41.087	1347.501	93.205	.421	19.842	948.396	226.127	1742.826	514.502	459.080	54.067

West, Above H.S.

	$x(33k\ell)$		$x_b^*(33k\ell)$	
	White	Non-white	White	Non-white
Fail	0	1	0	0.158
Pass	421	19	421	19.842
	421	20	421	20.000 .

The associated Analysis of Information Table 12.13 follows.

TABLE 12.13

ANALYSIS OF INFORMATION TABLE

	Component Due to	Information	D.F.
a)	$x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$	$2I(x:x_a^*) = 21.819$	18
b)	$x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$	$2I(x_b^*:x_a^*) = 5.868$	1
	less $x(3311)$, $x(3312)$	$2I(x:x_b^*) = 15.951$	17

Note that the OUTLIER entry in the computer output for x_a^*, West, Above H.S., White, Fail is 5.199 which is less than 5.868 as it should be. Also, since $2I(x_b^*:x_a^*)/2I(x:x_a^*) = 5.868/21.819 = 0.27$, the outlier values account for 27% of the "unexplained variation" $2I(x:x_a^*)$.

The computer output for the revised estimate x_b^* yields for South, H.S., Non-white, Fail the OUTLIER entry 3.69. Accordingly we now get a new revised estimate $x_c^*(ijk\ell)$. The estimate $x_c^*(ijk\ell)$ is obtained by fitting the marginals $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$, as for x_a^* and x_b^* except that the values $x(3311)$, $x(3312)$, and $x(4221)$, $x(4222)$ are not included, that is, $x_c^*(3311) = x(3311)$, $x_c^*(3312) = x(3312)$, $x_c^*(4221) = x(4221)$, $x_c^*(4222) = x(4222)$. The values of the estimate $x_c^*(ijk\ell)$ are given in Table 12.14.

TABLE 12.14

$x_c^*(ijk\ell)$

	i	East						North					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	62.107	8.828	44.571	4.608	7.608	0.278	21.717	2.611	19.850	1.339	4.267	0.215
	P	943.893	134.172	1880.428	194.392	320.392	11.722	868.282	104.389	2203.149	148.661	472.733	23.785

	i	West						South					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	13.007	0.940	11.119	0.769	0	0.164	43.433	10.356	28.743	16	7.578	0.892
	P	567.993	41.060	1347.880	93.231	421	19.836	950.566	226.644	1746.257	508	459.080	54.108

In particular, note that for West, Above H.S., and South, H.S., the corresponding observed and $x_c^*(ijk\ell)$ estimates are

West, Above H.S.

$x(33k\ell)$

	White	Non-white
Fail	0	1
Pass	421	19
	421	20

South, H.S.

$x(42k\ell)$

	White	Non-white
Fail	34	16
Pass	1741	508
	1775	524

$x_c^*(33k\ell)$

	White	Non-white
Fail	0	0.164
Pass	421	19.836
	421	20.000

$x_c^*(42k\ell)$

	White	Non-white
Fail	28.743	16
Pass	1746.257	508
	1775.000	524

The associated Analysis of Information Table 12.15 follows.

TABLE 12.15

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
a) $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$	$2I(x:x_a^*) = 21.819$	18
b) $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$ less $x(3311)$, $x(3312)$	$2I(x_b^*:x_a^*) = 5.868$	1
	$2I(x:x_b^*) = 15.951$	17
c) $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$ less $x(3311)$, $x(3312)$, $x(4221)$, $x(4222)$	$2I(x_c^*:x_b^*) = 4.511$	1
	$2I(x:x_c^*) = 11.440$	16

Note that the measure of effect $2I(x_c^*:x_b^*) = 4.511$ is greater than the OUTLIER entry for South, H.S., Non-white, Fail, 3.691, as it should be. Also, since $2I(x_c^*:x_b^*)/2I(x:x_b^*) = 4.511/15.951 = 0.28$, the outlier values $x(4221)$, $x(4222)$ account for 28% of the "unexplained variation" $2I(x:x_b^*)$.

The log-odds for the estimates x_b^* and x_c^* are also given by the parametric representation

$$\ln \frac{x^*(ijk1)}{x^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} \tag{12.3}$$

similar to that for x_a^*. The values of the parameters corresponding to x_b^* and x_c^* are given in Table 12.7 and the ratio of odds and odds of failure in Tables 12.8, 12.9 and 12.10.

We note that the results for home of record West and North are better than those for home of record South and East, even accounting for the outlier values.

13. Zero Marginals

As may be noted from the analysis in Section 12, zero occurrences in cells of the observed contingency table present no special problem provided that no marginal entering into the fitting specification is zero. When the latter is the case, however, the interpretation may be distorted because of inflated degrees of freedom. A procedure to circumvent this problem is similar to that used for getting revised estimates when outliers are indicated. We shall present the procedure in terms of a specific example.

A four-way cross-classification of 16,723 marines based on the variables home of record, level of education, AFQT, and boot camp completion is given in Table 13.1. We denote the occurrences in the four-way observed cross-classification or contingency table by $x(ijk\ell)$ with the notation

TABLE 13.1

x(ijkℓ)

i	EAST															
j		Below H.S.					H.S.					Above H.S.				
k		I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ	F	1	2	5	11	14	1	4	12	10	7	3	1	1	0	1
	P	19	50	343	364	302	339	568	959	395	328	191	126	119	11	14

i	NORTH															
j		Below H.S.					H.S.					Above H.S.				
k		I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ	F	0	1	4	1	1	0	2	2	2	4	1	0	0	0	1
	P	40	86	369	204	180	552	782	1088	264	252	347	188	146	17	18

i	WEST															
j		Below H.S.					H.S.					Above H.S.				
k		I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ	F	2	0	1	2	2	0	0	0	0	0	0	1	0	0	0
	P	49	27	202	146	108	386	329	492	226	219	365	125	115	28	15

i	SOUTH															
j		Below H.S.					H.S.					Above H.S.				
k		I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ	F	0	4	10	12	10	5	3	15	5	16	2	0	0	1	0
	P	26	71	528	474	409	270	352	973	516	747	230	144	195	57	55

Variable	Index	1	2	3	4	5
Home of Record	i	East	North	West	South	
Level of Education	j	Below H.S.	H.S.	Above H.S.		
AFQT	k	I	II	III	IV A	IV B
Boot Camp Completion	ℓ	Fail	Pass			

As in the analysis in Section 12, we are interested in the possible relationship of the variable fail or pass, as a dependent variable, on the independent or explanatory variables home of record, level of education and AFQT.

We summarize the results of fitting a sequence of nested marginals in the truncated Analysis of Information Table 13.2.

TABLE 13.2

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
a) $x(ijk.), x(...\ell)$	$2I(x:x_a^*) = 182.828$	59
e) $x(ijk.), x(i..\ell), x(..k\ell), x(.j.\ell)$	$2I(x_e^*:x_a^*) = 119.182$	9
	$2I(x:x_e^*) = 63.646$	50
n) $x(ijk.), x(..k\ell), x(ij.\ell)$	$2I(x_n^*:x_e^*) = 16.268$	6
	$2I(x:x_n^*) = 47.378$	44

We note that $2I(x:x_a^*) = 182.828$, 59 degrees of freedom, with $x_a^*(ijk\ell) = x(ijk.)\ x(...\ell)/n$ rejects the null hypothesis that boot camp completion is independent of the joint variable (home of record, length of education, and AFQT).

The value of $2I(x:x_n^*) = 47.378$, 44 degrees of freedom, implies that x_n^* is a good estimate and the value $2I(x_n^*:x_e^*) = 16.268$, 6 degrees of freedom, implies that the marginal $x(ij.\ell)$ and its associated interaction parameter for boot camp completion with home of record and level of education is significant. The values of $x_n^*(ijk\ell)$ are given in Table 13.3.

TABLE 13.3

$x_n^*(ijk\ell)$

EAST

i	EAST														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ F	0.424	1.059	7.542	11.412	12.565	3.729	6.022	10.888	6.405	6.955	2.411	1.515	1.524	0.197	0.35[illegible]
ℓ P	19.576	50.941	340.459	363.588	303.435	336.270	565.978	960.112	398.595	328.045	191.589	125.485	118.476	10.803	14.64[illegible]

i	NORTH														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ F	0.245	0.510	2.331	1.811	2.103	1.677	2.285	3.386	1.169	1.483	0.942	0.488	0.404	0.067	0.09[illegible]
ℓ P	39.755	86.490	370.669	203.189	178.897	550.323	781.715	1086.615	264.831	254.517	347.058	187.512	145.596	16.933	18.90[illegible]

i	WEST														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ F	0.505	0.257	2.057	2.116	2.065	0.000	0.000	0.000	0.000	0.000	0.543	0.180	0.175	0.060	0.04[illegible]
ℓ P	50.495	26.743	200.943	145.884	107.935	386.000	329.000	492.000	226.000	219.000	364.458	125.821	114.825	27.940	14.95[illegible]

i	SOUTH														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV B
ℓ F	0.438	1.214	9.270	11.781	13.297	3.167	3.924	11.634	8.650	16.625	0.918	0.547	0.789	0.332	0.41[illegible]
ℓ P	25.562	73.786	528.730	474.219	405.703	271.832	351.075	976.367	512.350	746.375	231.082	143.453	194.211	57.668	54.58[illegible]

The log-odds of fail to pass are given by the parametric representation

$$(13.1) \qquad \ell n \frac{x_n^*(ijk1)}{x_n^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{k1}^{k\ell} + \tau_{ij1}^{ij\ell} .$$

We note that in Table 13.1 no failures were recorded for recruits with home of record West and level of education H.S. for all AFQT's, that is, the observations $x(32k1)$ for $k = 1,2,3,4,5$ are all zero. As a consequence, the marginal $x(32.1) = 0$, and the estimates $x_n^*(32k\ell)$ based on fitting the marginals $x(ijk.)$, $x(..k\ell)$, $x(ij.\ell)$ are equal to $x(32k\ell)$. This distorts the interpretation on the basis of degrees of freedom and significant interaction parameters.

We shall therefore follow a procedure somewhat similar to that for OUTLIERS adjusting for the zero marginal values. The adjusted procedure is to delete the observations $x(32k\ell)$ from the estimation procedure. The revised estimates are derived by fitting the marginals $x(ijk.)$, $x(..k\ell)$, $x(ij.\ell)$ excluding the cells with home of record West and level of education H.S., that is, the cells $(32k\ell)$ and using the observed values $x(32k\ell)$ as the estimates for those cells. The revised procedure yields the Analysis of Information Table 13.4.

TABLE 13.4

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
r) $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$, $x(..k\ell)$, excluding $x(32k\ell)$	$2I(x:x_r^*) = 51.534$	45
s) $x(ijk.)$, $x(..k\ell)$, $x(ij.\ell)$, excluding $x(32k\ell)$	$2I(x_s^*:x_r^*) = 4.153$	5
	$2I(x:x_s^*) = 47.381$	40

Note that $2I(x:x_r^*)$ has 45 degrees of freedom compared to 50 for $2I(x:x_e^*)$ and $2I(x:x_s^*)$ has 40 degrees of freedom compared to 44 for $2I(x:x_n^*)$.

We now see that $2I(x_s^*:x_r^*) = 4.153$, 5 degrees of freedom, implies that adding $x(ij.\ell)$ to the set of fitted marginals, or the associated interaction parameters for home of record by level of education by failure, are not significant and $2I(x:x_r^*) = 51.534$, 45 degrees of freedom , implies that x_r^* is an acceptable fit. The values of $x_r^*(ijk\ell)$ are given in Table 13.5.

The parametric representation of the log-odds of failure in boot camp using the estimate $x_r^*(ijk\ell)$ is given by

$$(13.2) \qquad \ell n \frac{x_r^*(ijk1)}{x_r^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{k1}^{k\ell} .$$

Thus the log-odds depend only on an overall average effect τ_1^{ℓ} and additive effects due to home of record $\tau_{i1}^{i\ell}$, level of education $\tau_{j1}^{j\ell}$, and AFQT $\tau_{k1}^{k\ell}$, with no interaction effects. The values of the parameters in the representation of the log-odds are

$$\begin{array}{llll}
\tau_1^{\ell} = -4.376837 & & \tau_{21}^{j\ell} = 0.481840 \\
\tau_{11}^{i\ell} = 0.145880 & & \tau_{11}^{k\ell} = -0.665526 \\
\tau_{21}^{i\ell} = -1.148652 & & \tau_{21}^{k\ell} = -0.712272 \\
\tau_{31}^{i\ell} = -0.759926 & & \tau_{31}^{k\ell} = -0.639670 \\
\tau_{11}^{j\ell} = 1.029758 & & \tau_{41}^{k\ell} = -0.289594 .
\end{array}$$

For convenience we tabulate the odds of failure (to 1000) in Tables 13.6 and 13.7. Note that the overall odds of failure for this data (excluding West, H.S.) are 183/14888 = .0123 or 12.

Within a given home of record and for the same level of education the results for AFQT I, II, and III are apparently the same, with increasing odds of failure respectively for AFQT IV A and IV B.

The results for home of record North and West are consistently better than those for home of record South and East.

TABLE 13.5

$x_r^*(ijk\ell)$

i	EAST														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV
ℓ F	0.401	1.018	7.316	11.091	12.362	4.064	6.529	11.908	7.013	7.704	1.439	0.899	0.913	0.118	0.2[illegible]
ℓ P	19.590	50.982	340.684	363.909	303.637	335.935	565.470	959.093	397.988	327.296	192.561	126.101	119.087	10.882	14.7[illegible]

i	NORTH														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV
ℓ F	0.228	0.474	2.182	1.698	1.997	1.824	2.473	3.696	1.278	1.641	0.711	0.367	0.306	0.051	0.0[illegible]
ℓ P	39.772	86.527	370.818	203.302	179.003	550.176	781.527	1086.305	264.721	254.359	347.288	187.633	145.694	16.949	18.9[illegible]

i	WEST														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV
ℓ F	0.428	0.216	1.747	1.801	1.781	0	0	0	0	0	1.099	0.362	0.355	0.123	[illegible]
ℓ P	50.572	26.784	201.254	146.199	108.219	386	329	492	226	219	363.901	125.638	114.645	27.877	14.9[illegible]

i	SOUTH														
j	Below H.S.					H.S.					Above H.S.				
k	I	II	III	IV A	IV B	I	II	III	IV A	IV B	I	II	III	IV A	IV
ℓ F	0.462	1.273	9.803	12.473	14.242	2.846	3.508	10.489	7.815	15.213	1.489	0.882	1.284	0.540	[illegible]
ℓ P	25.538	73.727	528.197	473.527	404.757	272.154	351.492	977.511	513.185	747.788	230.511	143.118	193.716	57.460	[illegible]

TABLE 13.6

ODDS OF FAILURE $x_r^*(ijk1)/x_r^*(ijk2)$ TO 1000

	East	North	West	South
AFQT I				
Below H.S.	21	6	8	18
H.S.	12	3	0	10
Above H.S.	7	2	3	6
AFQT II				
Below H.S.	20	5	8	17
H.S.	12	3	0	10
Above H.S.	7	2	3	6
AFQT III				
Below H.S.	21	6	9	19
H.S.	12	3	0	11
Above H.S.	8	2	3	7
AFQT IV A				
Below H.S.	30	8	12	26
H.S.	18	5	0	15
Above H.S.	11	3	4	9
AFQT IV B				
Below H.S.	41	11	16	35
H.S.	24	6	0	20
Above H.S.	15	4	6	13

TABLE 13.7

ODDS OF FAILURE $x_r^*(ijk1)/x_r^*(ijk2)$ TO 1000

North

	I	II	III	IV A	IV B
Below H.S.	6	5	6	8	11
H.S.	3	3	3	5	6
Above H.S.	2	2	2	3	4

West

	I	II	III	IV A	IV B
Below H.S.	8	8	9	12	16
H.S.	0	0	0	0	0
Above H.S.	3	3	3	4	6

South

	I	II	III	IV A	IV B
Below H.S.	18	17	19	26	35
H.S.	10	10	11	15	20
Above H.S.	6	6	7	9	13

East

	I	II	III	IV A	IV B
Below H.S.	21	20	21	30	41
H.S.	12	12	12	18	24
Above H.S.	7	7	8	11	15

14. Acknowledgment

Without the interest, support and encouragement of Professor Henry Solomon and Professor Herbert Solomon this paper would never have been prepared. Professor C. T. Ireland's collaboration in the theory and computer application has been invaluable. Technical assistance by Mr. A. Azari and Dr. Marian Fisher has been most helpful. Comments by and discussions with Professor James T. Bennett and Professor Sheldon E. Haber have contributed to a better understanding of the particular applications. The bibliography is essentially a compilation of those prepared by Dr. Marvin A. Kastenbaum and Dr. H. H. Ku covering the period through 1971 and permission to use their results is gratefully acknowledged.

15. Bibliography

The bibliography lists publications on contingency table analysis through 1972. Additional references to related topics may be found in the bibliographies contained in the books by D. R. Cox (1970) and H. O. Lancaster (1969). We will appreciate your calling our attention to possible references omitted from the bibliography.

REFERENCES

1972

ADAM, J. and ENKE, H. (1972). Analyse mehrdimensionaler kontingenztafeln mit hilfe des informationsmasses von Kullback. Biometrische Zeitschrift **14**, 5, pp. 305-323.

BERKSON, J. (1972). Minimum discrimination information, the "no interaction" problem, and the logistic function. Biometrics **28**, 2, pp. 443-468.

BRUNDEN, M. N. (1972). The analysis of non-independent 2x2 tables from 2xc tables using rank sums. Biometrics **28**, 2, pp. 603-606.

CAUSEY, B. D. (1972). Sensitivity of raked contingency table totals to changes in problem conditions. Ann. Math. Statist. **43**, 2, pp. 656-658.

COX, D. R. (1972). The analysis of multivariate binary data. Appl. Statist. **21**, 2, pp. 113-120.

DARROCH, J. N. and RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. Ann. Math. Statist. **43**, 5, pp. 1470-1480.

FIENBERG, S. E. (1972). The analysis of incomplete multiway contingency tables. Biometrics **28**, 1, pp. 177-202.

GAIL, M. H. (1972). Mixed quasi-independence models for categorical data. Biometrics **28**, 3, pp. 703-712.

GART, J. J. (1972). Interaction tests for 2xsxt contingency tables. Biometrika **59**, 2, pp. 309-316.

GOKHALE, D. V. (1972). Analysis of log-linear models. J. Roy. Statist. Soc. Ser. B **34**, 3, pp. 371-376.

GOODMAN, L. A. and KRUSKAL, W. H. (1972). Measures of association for cross-classifications, IV: simplification of asymptotic variances. J. Amer. Statist. Assoc. **67**, pp. 415-421.

GRIZZLE, J. E. and WILLIAMS, O. D. (1972). Log-linear models and tests of independence for contingency tables. Biometrics **28**, 1, pp. 137-156.

GRIZZLE, J. E. and WILLIAMS. O. D. (1972). Contingency tables having ordered response categories. J. Amer. Statist. Assoc. **67**, pp. 55-63.

KOCH. G. G., IMREY, P. B., and REINFURT, D. W. (1972). Linear model analysis of categorical data with incomplete response vectors. Biometrics **28**, 3, pp. 663-692.

NATHAN, G. (1972). Asymptotic power of tests for independence in contingency tables from stratified samples. J. Amer. Statist. Assoc. **67**, pp. 917-920.

VICTOR, N. (1972). Zur klassifizierung mehrdimensionaler kontingenztafeln. Biometrics **28**, 2, pp. 427-442.

1971

ALTHAM, P. M. E. (1971). Exact Bayesian analysis of the intraclass 2x2 table. Biometrika **58**, 3, pp. 679-680.

ALTHAM, P. M. E. (1971). The analysis of matched proportions. Biometrika **58**, 3, pp. 561-576.

BELLE, G. V. and CORNELL, R. G. (1971). Strengthening tests of symmetry in contingency tables. Biometrics **27**, pp. 1074-1078.

BISHOP, Y. M. M. (1971). Effects of collapsing multidimensional contingency tables. Biometrics **27**, pp. 545-562.

COHEN, J. E. (1971). Estimation and interaction in a censored 2x2x2 contingency table. Biometrics **27**, 379-386.

DEMPSTER, A. P. (1971). An overview of multivariate data analysis. Journal Multivariate Analysis **1**, 316-347.

FRYER, J. G. (1971). On the homogeneity of the marginal distributions of a multidimensional contingency table. J. Roy. Statist. Soc. Ser. A 134, pp. 368-371.

GART, J. J. (1971). On the ordering of contingency tables for significance tests. Technometrics 13, pp. 910-911.

GART, J. J. (1971). The comparison of proportions: a review of significance tests, confidence intervals, and adjustments for stratification. Rev. Inst. Internat. Statist. 29, pp. 148-169.

GOKHALE, D. V. (1971). An iterative procedure for analysing log-linear models. Biometrics 27, pp. 681-687.

GOODMAN, L. A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. J. Amer. Statist. Assoc. 66, pp. 339-344.

GOODMAN, L. A. (1971). Some multiplicative models for the analysis of cross-classified data. Proc. 6th Berkeley Symp., Berkeley and Los Angeles, University of California Press.

GOODMAN, L. A. (1971). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. Technometrics 13, pp. 33-61.

GRIZZLE, J. E. (1971). Multivariate logit analysis. Biometrics 27, pp. 1057-1062.

JOHNSON, W. D., and KOCH, G. G. (1971). A note on the weighted least squares analysis of the Ries-Smith contingency table data. Technometrics 13, pp. 438-447.

KOCH, G. G., IMREY, P. B., and REINFURT, D. W. (1971). Linear model analysis of categorical data with incomplete response vectors. Institute of Statistics Mimeo Series No. 790, University of North Carolina.

KOCH, G. G., JOHNSON, W. D., and TOLLEY, H. D. (1971). An application of linear models to analyze categorical data pertaining to the relationship between survival and extent of disease. Institute of Statistics Mimeo Series No. 770, University of North Carolina.

KOCH. G. G. and REINFURT, D. W. (1971). The analysis of categorical data from mixed models. Biometrics 27, pp. 157-173.

KU, H. H. (1971). Analysis of information - an alternative approach to the detection of a correlation between the sexes of adjacent sibs in human families. Biometrics 27, pp. 175-182.

KU, H. H., VARNER, R., and KULLBACK, S. (1971). On the analysis of multidimensional contingency tables. J. Amer. Statist. Assoc. 66, pp. 55-64.

KULLBACK, S. (1971). Marginal homogeneity of multidimensional contingency tables. Ann. Math. Statist. 42, pp. 594-606.

KULLBACK, S. (1971). The homogeneity of the sex ratio of adjacent sibs in human families. Biometrics 27, pp. 452-457.

NAM, J. (1971). On two tests for comparing matched proportions. Biometrics 27, pp. 945-959.

PEACOCK, P. B. (1971). The non-comparability of relative risks from different studies. Biometrics 27, pp. 903-907.

PERITZ, E. (1971). Estimating the ratio of two marginal probabilities in a contingency table. Biometrics 27, pp. 223-225.

RATCLIFF, D. (1971). Topics on independence and correlation for bounded sum variables. Ph.D. thesis, School of Mathematical Sciences, the Flinders University of South Australia, June 1971.

SIMON, G. A. (1971). Information distances and exponential families, with applications to contingency tables. Technical Report No. 32, November 26, 1971, Department of Statistics, Stanford University.

THOMAS, D. G. (1971). Exact confidence limits for the odds ratio in a 2x2 table. Appl. Statist. **20**, pp. 105-110.

YASAIMAIBODI (YASSAEE), HEDAYAT (1971). On comparison of various estimators and their associated statistics in rxc and rxcx2 contingency tables. Ph.D. dissertation, The George Washington University.

ZELEN, M. (1971). The analysis of several 2x2 contingency tables. Biometrika **58**, pp. 129-137.

1970

ALTHAM, PATRICIA M. E. (1970). The measurement of association of rows and columns for an rxs contingency table. J. Roy. Statist. Soc. Ser. B **32**, pp. 63-73.

BHAPKAR, V. P. (1970). Categorical data analysis of some multivariate tests. Essays in Probability and Statistics (R. C. Bose et al., eds.). The University of North Carolina Press, pp. 85-110.

CAMPBELL, L. L. (1970). Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities. Ann. Math. Statist. **41**, pp. 1011-1015.

COX, D. R. (1970). The Analysis of Binary Data. Methuen & Co., Ltd., London.

CRADDOCK, J. M. and FLOOD, C. R. (1970). The distribution of the chi-square statistic in small contingency tables. Appl. Statist. **19**, pp. 173-181.

FIENBERG, S. E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables. J. Amer. Statist. Assoc. **65**, 332, pp. 1610-1616.

FIENBERG, S. E. (1970). An iterative procedure for estimation in contingency tables. Ann. Math. Statist. **41**, pp. 907-917.

FIENBERG, S. E. (1970). The analysis of multidimensional contingency tables. Ecology **51**, 2, pp. 419-433.

FIENBERG, S. E. and GILBERT, J. P. (1970). Geometry of a two by two contingency table. J. Amer. Statist. Assoc. **65**, pp. 694-701.

FIENBERG, S. E. and HOLLAND, P. W. (1970). Methods for eliminating zero counts in contingency tables. Random Counts in Scientific Work (G. P. Patil, ed.). The Pennsylvania State University Press.

GOOD, I. J., GOVER, T. N., and MITCHELL, G. J. (1970). Exact distributions for χ-squared and for the likelihood-ratio statistic for the equiprobable multinomial distribution. J. Amer. Statist. Assoc. **65**, pp. 267-283.

GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications. J. Amer. Statist. Assoc. **65**, pp. 226-256.

KASTENBAUM, M. A. (1970). A review of contingency tables. Essays in Probability and Statistics (R. C. Bose et al., eds.). The University of North Carolina Press, pp. 407-438.

KULLBACK, S. (1970). Various applications of minimum discrimination information estimation, particularly to problems of contingency table analysis. Proceedings of the Meeting on Information Measures, University of Waterloo, Ontario, Canada, April 10-14, 1970, I-33-I-66.

KULLBACK, S. (1970). Minimum discrimination information estimation and application. Proceedings of the Sixteenth Conference on the Design of Experiments in Army Research, Development and Testing, 21 October 1970. ARO-D Report 71-3, 1-38 Proceedings of the Conference.

MANTEL, N. (1970). Incomplete contingency tables. Biometrics **26**, pp. 291-304.

MOLK, YEHUDA (1970). On estimation of probabilities in contingency tables with restrictions on marginals. Ph.D. dissertation, The George Washington University, February 1970.

ODOROFF, C. L. (1970). Minimum logit chi-square estimation and maximum likelihood estimation in contingency tables. J. Amer. Statist. Assoc. 65, 332, pp. 1617-1631.

WAGNER, S. S. (1970). The maximum-likelihood estimate for contingency tables with zero diagonal. J. Amer. Statist. Assoc. 65, 331, pp. 1362-1383.

1969

ALTHAM, PATRICIA M. E. (1969). Exact Bayesian analysis of a 2x2 contingency table and Fisher's "exact" significance test. J. Roy. Statist. Soc. Ser. B. 31, pp. 261-269.

ARGENTIERO, P. D. (1969). χ-squared statistic for goodness of fit test, its derivation and tables. NASA Technical Report, TR-R-313.

BISHOP, Y. M. M. (1969). Full contingency tables, logits, and split contingency tables. Biometrics 25, pp. 383-400.

BISHOP, Y. M. M. and FIENBERG, S. E. (1969). Incomplete two-dimensional contingency tables. Biometrics 25, pp. 119-128.

DEMPSTER, A. P. (1969). Some theory related to fitting exponential models. Research Report S-4, Department of Statistics, Harvard University.

FIENBERG, S. E. (1969). Preliminary graphical analysis and quasi-independence for two-way contingency tables. Appl. Statist. 18, pp. 153-168.

GOODMAN, L. A. (1969). On partition χ-squared and detecting partial association in the three-way contingency tables. J. Roy. Statist. Soc. Ser. B. 31, pp. 486-498.

GRIZZLE, J. E., STARMER, C. F., and KOCH, G. G. (1969). Analysis of categorical data by linear models. Biometrics 25, pp. 489-504.

HEALY, M. J. R. (1969). Exact tests of significance in contingency tables. Technometrics **11**, pp. 393-395.

IRELAND, C. T., KU, H. H., and KULLBACK, S. (1969). Symmetry and marginal homogeneity of an rxr contingency table. J. Amer. Statist. Assoc. **64**, pp. 1323-1341.

KOCH, G. G. (1969). The effect of non-sampling errors on measures of association in 2x2 contingency tables. J. Amer. Statist. Assoc. **64**, pp. 852-863.

KU, H. H. and KULLBACK, S. (1969). Analysis of multidimensional contingency tables: an information theoretical approach. Contributed papers, 37th Session of the International Statistical Institute, pp. 156-158.

LANCASTER, H. O. (1969). Contingency tables of higher dimensions. Bulletin of the International Statistical Institute. **43**, 1, pp. 143-151.

LANCASTER, H. O. (1969). The Chi-Squared Distribution. Wiley, New York.

NAGNUR, B. N. (1969). LAMST and the hypotheses of no three factor interaction in contingency tables. J. Amer. Statist. Assoc. **64**, pp. 207-215.

PLACKETT, R. L. (1969). Multidimensional contingency tables. A Survey of Models and Methods, Bulletin of the International Statistical Institute. **43**, 1, pp. 133-142.

1968

BENNETT, B. M. (1968). Notes on χ-squared tests for matched samples. J. Roy.Statist. Soc. Ser. B **30**, pp. 368-370.

BERKSON, J. (1968). Application of minimum logit chi-squared estimate to a problem of Grizzle with a notation on the problem of no interaction. Biometrics **24**, pp. 75-96.

BHAPKAR, V. P. (1968). On the analysis of contingency tables with a quantitative response. Biometrics **24**, pp. 329-338.

BHAPKAR, V. P. and KOCH, G. G. (1968). Hypotheses of "no interaction" in multidimensional contingency tables. Technometrics **10**, pp. 107-123.

BHAPKAR, V. P. and KOCH, G. G. (1968). On the hypotheses of "no interaction" in contingency tables. Biometrics **24**, pp. 567-594.

FIENBERG, S. E. (1968). The geometry of an rxc contingency table. Ann. Math. Statist. **39**, pp. 1186-1190.

GOODMAN, L. A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. J. Amer. Statist. Assoc. **63**, pp. 1091-1131.

HAMDAN, M. A. (1968). Optimum choice of classes for contingency tables. J. Amer. Statist. Assoc. **63**, pp. 291-297.

IRELAND, C. T. and KULLBACK, S. (1968). Contingency tables with given marginals. Biometrika **55**, pp. 179-188.

IRELAND, C. T. and KULLBACK, S. (1968). Minimum discrimination information estimation. Biometrics **24**, pp. 707-713.

KU, H. H. and KULLBACK, S. (1968). Interaction in multidimensional contingency tables: an information theoretic approach. J. Res. Nat. Bur. Standards Sect. B **72**, pp. 159-199.

KU, H. H., VARNER, R., and KULLBACK, S. (1968). Analysis of multidimensional contingency tables. Proceedings of the Fourteenth Conference on the Design of Experiments in Army Research, Development and Testing. ARO-D Report 69-2.

KULLBACK, S. (1968). Information Theory and Statistics. Dover Pub., Inc., New York.

KULLBACK, S. (1968). Probability densities with given marginals. Ann. Math. Statist. 39, pp. 1236-1243.

LEYTON, M. K. (1968). Rapid calculation of exact probabilities for 2x3 contingency tables. Biometrics 24, pp. 714-717.

MATHIEU, JEAN-RENE and LAMBERT, E. (1968). Un test de l'identite des marges dun tableau de correlation. C. R. Acad. Sci. Paris 267, pp. 832-834.

MOSTELLER, F. (1968). Association and estimation in contingency tables. J. Amer. Statist. Assoc. 63, pp. 1-28.

SLAKTER, M. J. (1968). Accuracy of an approximation to the power of the chi-square goodness of fit test with small but equal expected frequencies. J. Amer. Statist. Assoc. 63, pp. 912-924.

SUGIURA, N. and OTAKE, M. (1968). Numerical comparison of improvised methods of testing in contingency tables with small frequencies. Ann. Math. Statist. 20, pp. 507-517.

1967

BENNETT, B. M. (1967). Tests of hypothesis concerning matched samples. J. Roy. Statist. Soc. Ser. B 29, pp. 468-474.

BISHOP, Y. M. M. (1967). Multidimensional contingency tables: cell estimates. Ph.D. dissertation, Harvard University.

BLOCH, D. A. and WATSON, G. S. (1967). A Bayesian study of the multinomial distribution. Ann. Math. Statist. 38, pp. 1423-1435.

COX, D. R. and LAUH, E. (1967). A note on the graphical analysis of multidimensional contingency tables. Technometrics 9, pp. 481-488.

GOOD, I. J. (1967). A Bayesian significance test for multinomial distributions. J. Roy. Statist. Soc. Ser. B 29, pp. 339-431.

1966

ARMITAGE, P. (1966). The chi-square test for heterogeneity of proportions after adjustment for stratification. J. Roy. Statist. Soc. Ser. B 28, pp. 150-163.

BHAPKAR, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. J. Amer. Statist. Assoc. 61, pp. 228-235.

BHAPKAR, V. P. (1966). Notes on analysis of categorical data. Institute of Statistics Mimeo Series No. 477, University of North Carolina.

BHAT, B. R. and KULKARNI, S. R. (1966). LAMP test of linear and loglinear hypotheses in multinomial experiments. J. Amer. Statist. Assoc. 61, pp. 236-245.

COX, D. R. (1966). A simple example of a comparison involving quantal data. Biometrika 53, pp. 215-220.

CRADDOCK, J. M. (1966). Testing the significance of a 3x3 contingency table. The Statistician 16, pp. 87-94.

GABRIEL, K. R. (1966). Simultaneous test procedures for multiple comparison on categorical data. J. Amer. Statist. Assoc. 61, pp. 1081-1096.

GART, J. J. (1966). Alternative analyses of contingency tables. J. Roy. Statist. Soc. Ser. B 28, pp. 164-179.

GOOD, I. J. (1966). How to estimate probabilities. J. Inst. Math. Appl. 2, pp. 364-383.

KULLBACK, S. and KHAIRAT, M. A. (1966). A note on minimum discrimination information. Ann. Math. Statist. 37, pp. 279-280.

MANTEL, N. (1966). Models for complex contingency tables and polychotomous dosage response curves. Biometrics 22, pp. 83-95.

1965

ASANO, C. (1965). On estimating multinomial probabilities by pooling incomplete samples. Ann. Inst. Statist. Math. **17**, pp. 1-14.

BHAPKAR, V. P. and KOCH, G. G. (1965). On the hypothesis of "no interaction" in three-dimensional contingency tables. Institute of Statistics Mimeo Series No. 440, University of North Carolina.

BHAPKAR, V. P. and KOCH, G. G. (1965). Hypothesis of no interaction in four-dimensional contingency tables. Institute of Statistics Mimeo Series No. 449, University of North Carolina.

BHAT, B. R. and NAGNUR, B. N. (1965). Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypotheses. Biometrika **52**, 3 and 4, pp. 459-468.

BIRCH, M. W. (1965). The detection of partial association II: the general case. J. Roy. Statist. Soc. Ser. B **27**, pp. 111-124.

CAUSSINUS, H. (1965). Contribution a l'analyse statistique des tableaux de correlation. Ann. Fac. Sci. Univ. Toulouse **29**, pp. 77-182.

GOOD, I. J. (1965). The Estimation of Probabilities: An Essay on Modern Bayesian Methods. Research Monograph, 30. The MIT Press, Cambridge, Massachusetts.

KASTENBAUM, M. A. (1965). Contingency tables: a review. MRC Technical Summary Report No. 596. Mathematical Research Center, The University of Wisconsin.

KATTI, S. K. and SASTRY, A. N. (1965). Biological examples of small expected frequencies and the chi-square test. Biometrics **21**, pp. 49-54.

LANCASTER, H. O. and BROWN, T. A. I. (1965). Size of χ-squared test in the symmetrical multinomials. Austral. J. Statist. **7**, p. 40.

LEWONTIN, R. C. and FELSENSTEIN (1965). The robustness of homogeneity tests in 2xn tables. Biometrics **21**, pp. 19-33.

MOTE, V. L. and ANDERSON, R. L. (1965). An investigation of the effect of misclassification on the properties of chi-squared tests in the analysis of categorical data. Biometrika **52**, pp. 95-109.

RADHAKRISHNA, S. (1965). Combination of results from several 2x2 contingency tables. Biometrics **21**, pp. 86-98.

1964

ALLISON, H. E. (1964). Computational forms for chi-square. Amer. Statist. **18**, 1, pp. 17-18.

BENNETT, B. M. and NAKAMURA, E. (1964). Tables for testing significance in a 2x3 contingency table. Technometrics **6**, 4, pp. 439-458.

BROSS, I. D. J. (1964). Taking a covariable into account. J. Amer. Statist. Assoc. **59**, 307, pp. 725-736.

CHEW, V. (1964). Application of the negative binomial distribution with probability of misclassification. Virginia Journal of Science **15**, 1, pp. 34-40.

GOODMAN, L. A. (1964). Simultaneous confidence limits for cross-product ratios in contingency tables. J. Roy. Statist. Soc. Ser. B **26**, 1, pp. 86-102.

GOODMAN, L. A. (1964). Simple methods for analyzing three-factor interaction in contingency tables. J. Roy. Statist. Soc. **59**, pp. 319-352.

GOODMAN, L. A. (1964). Interactions in multidimensional contingency tables. Ann. Math. Statist. **35**, 2, pp. 632-646.

GOODMAN, L. A. (1964). Simultaneous confidence intervals for contrasts among multinomial populations. Ann. Math. Statist. **35**, 2, pp. 716-725.

HARKNESS, W. L. and KATZ, L. (1964). Comparison of the power functions for the test of independence in 2x2 contingency tables. Ann. Math. Statist. **35**, 3, pp. 1115-1127.

LINDLEY, D. V. (1964). The Bayesian analysis of contingency tables. Ann. Math. Statist. **35**, 4, pp. 1622-1643.

PLACKETT, R. L. (1964). The continuity correction in 2x2 tables. Biometrika **21**, Parts 3 and 4, pp. 327-338.

PUTTER, J. (1964). The χ^2 goodness-of-fit test for a class of cases of dependent observations. Biometrika **51**, pp. 250-252.

SOMERS, R. H. (1964). Simple measures of association for the triple dichotomy. J. Roy. Statist. Soc. Ser. A **127**, 3, pp. 409-415.

TALLIS, G. M. (1964). The use of models in the analysis of some classes of contingency tables. Biometrics **24**, 4, pp. 832-839.

1963

BENNETT, B. M. and NAKAMURA, E. (1963). Tables for testing significance in a 2x3 contingency table. Technometrics **5**, 4, pp. 501-511.

BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. J. Roy. Statist. Soc. Ser. B **25**, 1, pp. 220-233.

DARROCH, J. N. and SILVEY, S. D. (1963). On testing more than one hypothesis. Ann. Math. Statist. **34**, 2, pp. 555-567.

DIAMOND, E. L. (1963). The limiting power of categorical data chi-square tests analogous to normal analysis of variance. Ann. Math. Statist. **34**, 4, pp. 1432-1441.

EDWARDS, A. W. F. (1963). The measure of association in a 2x2 table. J. Roy. Statist. Soc. Ser. A **126**, 1, pp. 109-114.

FELDMAN, S. E. and KLINGER, E. (1963). Short cut calculation of the Fisher-Yates exact test. Psychometrika **28**, 3, pp. 289-291.

GOLD, R. A. (1963). Tests auxiliary to χ^2 tests in a Markov chain. Ann. Math. Statist. 34, 1, pp. 56-74.

GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. Ann. Math. Statist. 34, 3, pp. 911-934.

GOODMAN, L. A. (1963). On methods for comparing contingency tables. J. Roy. Statist. Soc. Ser. A 126, 1, pp. 94-108.

GOODMAN, L. A. (1963). On Plackett's test for contingency table interactions. J. Roy. Statist. Soc. Ser. B 25, 1, pp. 179-188.

GOODMAN, L. A. and KRUSKAL, W. H. (1963). Measures of association for cross classification III: approximate sampling theory. J. Amer. Statist. Assoc. 58, pp. 310-364.

KU, H. H. (1963). A note on contingency tables involving zero frequencies and the 2I test. Technometrics 5, 3, pp. 398-400.

MANTEL, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. J. Amer. Statist. Assoc. 58, pp. 690-700.

NEWELL, D. J. (1963). Misclassification in 2x2 tables. Biometrics 19, 1, pp. 187-188.

OKAMATO, M. (1963). Chi-square statistic based on the pooled frequencies of several observations. Biometrika 50, pp. 524-528.

RIES, P. N. and SMITH, H. (1963). The use of chi-square for preference testing in multidimensional problems. Chem. Eng. Prog. Symposium Series 59, 42, pp. 39-43.

WALSH, J. E. (1963). Loss in test efficiency due to misclassification for 2x2 tables. Biometrics 19, 1, pp. 158-162.

1962

DALY, C. (1962). A simple test for trends in a contingency table. Biometrics 18, 1, pp. 114-119.

DARROCH, J. N. (1962). Interactions in multi-factor contingency tables. J. Roy. Statist. Soc. Ser. B 24, 1, pp. 251-263.

FISHER, SIR RONALD A. (1962). Confidence limits for a cross-product ratio. Austral. J. Statist. 4, 1, p. 41.

GART, J. J. (1962). Approximate confidence limits for relative risks. J. Roy. Statist. Soc. Ser. B 24, 2, pp. 454-463.

GART, J. J. (1962). On the combination of relative risks. Biometrics 18, 4, pp. 601-610.

KINCAID, W. M. (1962). The combination of 2xm contingency tables. Biometrics 18, 2, pp. 224-228.

KULLBACK, S. KUPPERMAN, M. and KU, H. H. (1962). An application of information theory to the analysis of contingency tables with a table of 2N ln N, N = 1(1)10,000. J. Res. Nat. Bur. Standards Sect. B 66, pp. 217-243.

KULLBACK, S., KUPPERMAN, M., and KU, H. H. (1962). Tests for contingency tables and Markov chains. Technometrics 4, 4, pp. 573-608.

LEWIS, B. N. (1962). On the analysis of interaction in multi-dimensional contingency tables. J. Roy. Statist. Soc. Ser. A 125, 1, pp. 88-117.

PLACKETT, R. L. (1962). A note on interactions in contingency tables. J. Roy. Statist. Soc. Ser. B 24, 1, pp. 162-166.

TALLIS, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. Biometrics 18, 3, pp. 342-353.

1961

BERGER, A. (1961). On comparing intensities of association between two binary characteristics in two different populations. J. Amer. Statist. Assoc. 56, pp. 889-908.

BHAPKAR, V. P. (1961). Some tests for categorical data. Ann. Math. Statist. 32, 1, pp. 72-83.

BILLINGSLEY, P. (1961). Statistical Inference for Markov Processes. Statistical Research Monographs, 2, The University of Chicago Press.

CLARINGBOLD, P. J. (1961). The use of orthogonal polynomials in the partition of chi-square. Austral. J. Statist. 3, 2, pp. 48-63.

FRIEDLANDER, D. (1961). A technique for estimating a contingency table, given the marginal totals and some supplementary data. J. Roy. Statist. Soc. Ser. A 124, 3, pp. 412-420.

GREGORY, G. (1961). Contingency tables with a dependent classification. Austral. J. Statist. 3, 2, pp. 42-47.

GRIZZLE, J. E. (1961). A new method of testing hypotheses and estimating parameters for the logistic model. Biometrics 17, 3, pp. 372-385.

KENDALL, M. G. and STUART, A. (1961). The Advanced Theory of Statistics. 2, Charles Griffin and Company, London.

OKAMATO, M. and ISHII, G. (1961). Test of independence in intraclass 2x2 tables. Biometrika 48, pp. 181-190.

ROGOT, E. (1961). A note on measurement errors and detecting real differences. J. Amer. Statist. Assoc. 56, pp. 314-319.

SCHULL, W. J. (1961). Some problems of analysis of multi-factor tables. Bull. Inst. Internat. Statist. 28, Part 3, pp. 259-270.

YATES, F. (1961). Marginal percentages in multiway tables of quantal data with disproportionate frequencies. Biometrics 17, 1, pp. 1-9.

1960

BENNETT, B. M. and HSU, P. (1960). On the power function of the exact test for the 2x2 contingency table. Biometrika 47, pp. 393-398.

GRIDGEMAN, N. T. (1960). Card-matching experiments: a conspectus of theory. J. Roy. Statist. Soc. Ser. A 123, 1, pp. 45-49.

ISHII, G. (1960). Intraclass contingency tables. Ann. Inst. Statist. Math. 12, pp. 161-207; corrections, p. 279.

KASTENBAUM, M. A. (1960). A note on the additive partitioning of chi-square in contingency tables. Biometrics 16, 3, pp. 416-422.

KUPPERMAN, M. (1960). On comparing two observed frequency counts. Appl. Statist. 9, 1, pp. 37-42.

LANCASTER, H. O. (1960). On tests on independence in several dimensions. J. Austral. Math. Soc. 1, pp. 241-254.

ROBERTSON, W. H. (1960). Programming Fisher's exact method of comparing two percentages. Technometrics 2, 1, pp. 103-107.

1959

ANDERSON, R. L. (1959). Use of contingency tables in the analysis of consumer preference studies. Biometrics 15, 4, pp. 582-590.

CHAKRAVARTI, I. M. and RAO, C. R. (1959). Tables for some small sample tests of significance for Poisson distributions and 2x3 contingency tables. Sankhya 21, Parts 3 and 4, pp. 315-326.

GOODMAN, L. A. and KRUSKAL, W. H. (1959). Measures of association for cross classification II: further discussion and references. J. Amer. Statist. Assoc. 54, pp. 123-163.

HALDANE, J. B. S. (1959). The analysis of heterogeneity, I. Sankhya 21, Parts 3 and 4, pp. 209-216.

HOYT, C. J., KRISHNAIAH, P. R., and TORRANCE, E. P. (1959). Analysis of complex contingency data. Journal of Experimental Education 27, pp. 187-194.

KASTENBAUM, M. A. and LAMPHIEAR, D. E. (1959). Calculation of chi-square to test the no three-factor interaction hypothesis. Biometrics 15, 1, pp. 107-115.

KULLBACK, S. (1959). Information Theory and Statistics. John Wiley and Sons, New York.

KUPPERMAN, M. (1959). A rapid significance test for contingency tables. Biometrics 15, 4, pp. 625-628.

NASS, C. A. G. (1959). The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. Biometrika 46, pp. 365-385.

SILVEY, S. D. (1959). The Lagrangian multiplier test. Ann. Math. Statist. 30, 2, pp. 389-407.

SOMERS, R. H. (1959). The rank analogue of product-moment partial correlation and regression, with application to manifold, ordered contingency tables. Biometrika 46, pp. 241-246.

STEYN, H. S. (1959). On χ^2-tests for contingency tables of negative binomial type. Statistica Neerlandica 13, pp. 433-444.

WEINER, I. B. (1959). A note of the use of Mood's likelihood ratio test for item analyses involving 2x2 tables with small samples. Psychometrika 24, 4, pp. 371-372.

1958

BLALOCK. H. M., Jr. (1958). Probabilistic interpretations for the mean square contingency. J. Amer. Statist. Assoc. 53, pp. 102-105.

KASTENBAUM, M. A. (1958). Estimation of relative frequencies of four sperm types in Drosophila melanogaster. Biometrics 14, 2, pp. 223-228.

MITRA, S. K. (1958). On the limiting power function of the frequency chi-square test. Ann. Math. Statist. 29, pp. 1221-1233.

SNEDECOR, G. W. (1958). Chi-square of Bartlett, Mood and Lancaster in a 2^3 contingency table. Biometrics 14, 4, pp. 560-562 (Query).

1957

BROSS, I. D. J. and KASTEN, E. L. (1957). Rapid analysis of 2x2 tables. J. Amer. Statist. Assoc. 52, pp. 18-28.

CORSTEN, L. C. A. (1957). Partition of experimental vectors connected with multinomial distributions. Biometrics 13, 4, pp. 451-484.

EDWARDS, J. H. (1957). A note on the practical interpretation of 2x2 tables. Brit. J. Prev. Soc. Med. 11, pp. 73-78.

LANCASTER, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. Biometrika 44. pp. 289-292.

MOTE, V. L. (1957). An investigation of the effect of misclassification of the chi-square tests in the analysis of categorical data. Unpublished Ph.D. dissertation, North Carolina State College, Raleigh, North Carolina (also Institute of Statistics Mimeo Series No. 182).

ROY, S. N. (1957). Some Aspects of Multivariate Analysis. John Wiley and Sons, New York.

SAKODA, J. M. and COHEN, B. H. (1957). Exact probabilities for contingency tables using binomial coefficients. Psychometrika 22, 1, pp. 83-86.

WOOLF, B. (1957). The log likelihood ratio test (the G-test). Methods and tables for tests of heterogeneity in contingency tables. Annals of Human Genetics 21, pp. 397-409.

1956

FISHMAN, J. A. (1956). A note on Jenkins' "Improved Method for Tetrachoric r." _Psychometrika_ **20**, 3, pp. 305.

GOOD, I. J. (1956). On the estimation of small frequencies in contingency tables. _J. Roy. Statist. Soc. Ser. B_ **18**, 1, pp. 113-124.

GRIDGEMAN, N. T. (1956). A tasting experiment. _Appl. Statist._ **5**, 2, pp. 106-112.

LEANDER, E. K. and FINNEY, D. J. (1956). An extension of the use of the χ^2 test. _Appl. Statist._ **5**, 2, pp. 132-136.

MAINLAND, D., HERRERA, L. and SUTCLIFFE, M. I. (1956). Statistical tables for use with binomial samples - contingency tests, confidence limits, and sample size estimates. New York University College of Medicine, New York.

ROY, S. N. and KASTENBAUM, M. A. (1956). On the hypothesis of no "interaction" in a multiway contingency table . _Ann. Math. Statist._ **27**, 3, pp. 749-757.

ROY, S. N. and MITRA, S. K. (1956). An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. _Biometrika_ **43**, Parts 3 and 4, pp. 361-376.

WATSON, G. S. (1956). Missing and "mixed-up" frequencies in contingency tables. _Biometrics_ **12**, 1, pp. 47-50.

1955

ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. _Biometrics_ **11**, 3, pp. 375-386.

ARMSEN, P. (1955). Tables for significance tests of 2x2 contingency tables. _Biometrika_ **42**, pp. 494-505.

COCHRAN, W. G. (1955). A test of a linear function of the deviations between observed and expected numbers. _J. Amer. Statist. Assoc._ **50**, pp. 377-397.

HALDANE, J. B. S. (1955). Substitutes for χ^2. Biometrika 42, pp. 265-266.

HALDANE, J. B. S. (1955). A problem in the significance of small numbers. Biometrika 42, pp. 266-267.

HALDANE, J. B. S. (1955). The rapid calculation of χ^2 as a test of homogeneity from a 2xn table. Biometrika 42, pp. 519-520.

JENKINS, W. L. (1955). An improved method for tetrachoric r. Psychometrika 20, 3, pp. 253-258.

KASTENBAUM, M. A. (1955). Analysis of data in multiway contingency tables. Unpublished doctoral dissertation. North Carolina State College, October 1955.

LESLIE, P. H. (1955). A simple method of calculating the exact probability in 2x2 contingency tables with small marginal totals. Biometrika 42, pp. 522-523.

MITRA, S. K. (1955). Contributions to the statistical analysis of categorical data. North Carolina Institute of Statistics Mimeograph Series No. 142, December 1955.

ROY, S. N. and KASTENBAUM, M. A. (1955). A generalization of analysis of variance and multivariate analysis to data based on frequencies in qualitative categorical or class intervals. North Carolina Institute of Statistics Mimeograph Series No. 131, June 1955.

ROY, S. N. and MITRA, S. K. (1955). An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. North Carolina Institute of Statistics Mimeograph Series No. 139, November 1955.

SEKAR, C. C., AGARIVALA, S. P. and CHAKRABORTY, P. N. (1955). On the power function of a test of significance for the difference between two proportions. Sankhya 15, Part 4, pp. 381-390.

STUART, A. (1955). A test of homogeneity of the marginal distributions in a two-way classification. Biometrika **42**, pp. 412-416.

WOOLF, B. (1955). On estimating the relation between blood group and disease. Annals of Human Genetics **19**, pp. 251-253.

YATES, F. (1955). A note on the application of the combination of probabilities test to a set of 2x2 tables. Biometrika **42**, pp. 401-411.

YATES, F. (1955). The use of transformations and maximum likelihood in the analysis of quantal experiments involving two treatments. Biometrika **42**, pp. 382-403.

1954

BROSS, I. D. J. (1954). Misclassification in 2x2 tables. Biometrics **10**, 4, pp. 478-486.

COCHRAN, W. G. (1954). Some methods for strengthening the common chi-square tests. Biometrics **10**, 4, pp. 417-451.

DAWSON, R. B. (1954). A simplified expression for the variance of the χ^2 function on a contingency table. Biometrika **41**, p. 280.

GOODMAN, L. A. and KRUSKAL, W. H. (1954). Measures of association for cross classification. J. Amer. Statist. Assoc. **49**, pp. 732-764.

KIMBALL, A. W. (1954). Short-cut formulas for the exact partition of chi-square in contingency tables. Biometrics **10**, 4, pp. 452-458.

MCGILL, W. J. (1954). Multivariate information transmission. Psychometrika **19**, 2, pp. 97-116.

1952

COCHRAN, W. G. (1952). The χ^2 test of goodness of fit. Ann. Math. Statist. **23**, 3, pp. 315-345.

DYKE, G. V. and PATTERSON, H. D. (1952). Analysis of factorial arrangements when the data are proportions. Biometrics **8**, pp. 1-12.

1951

LANCASTER, H. O. (1951). Complex contingency tables treated by the partition of chi-square. J. Roy. Statist. Soc. Ser. B **13**, pp. 242-249.

SIMPSON, C. H. (1951). The interpretation of interaction in contingency tables. J. Roy. Statist. Soc. Ser. B **13**, pp. 238-241.

1950

TOCHER, K. D. (1950). Extension of the Neyman-Pearson theory of tests to discontinuous variates. Biometrika **37**, pp. 130-144.

1949

HSU, P. L. (1949). The limiting distributions of functions of sample means and application to testing hypotheses. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability (1945, 1946), University of California Press, Berkeley and Los Angeles.

IRWIN, J. O. (1949). A note on the subdivision of chi-square into components. Biometrika **36**, pp. 130-134.

LANCASTER, H. O. (1949). The derivation and partition of chi-square in certain discrete distributions. Biometrika **36**, pp. 117-129.

1948

YATES, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. Biometrika **35**, pp. 176-181.

1947

BARNARD, G. A. (1947). Significance tests for 2x2 tables. Biometrika **34**, pp. 123-138.

PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table. Biometrika 34, pp 139-167.

1946

CRAMER, H. (1946). Mathematical Methods of Statistics. Princeton University Press, p. 424.

1945

NORTON, H. W. (1945). Calculation of chi-square for complex contingency tables. J. Amer. Statist. Assoc. 40, pp. 251-258.

1937

HALDANE, J. B. S. (1937). The exact value of the moments of the distribution of χ^2 used as a test of goodness of fit, when expectations are small. Biometrika 29, pp. 133-143.

1935

BARTLETT, M. S. (1935). Contingency table interactions. J. Roy. Statist. Soc. Supplement 2, pp. 248-252.

WILKS, S. S. (1935). The likelihood test of independence in contingency tables. Ann. Math. Statist. 6, pp. 190-196.

1934

FISHER, R. A. (1934). Statistical Methods for Research Workers, Edinburgh 5th and subsequent editions. Oliver and Boyd Ltd., Section 21.02.

YATES, F. (1934). Contingency tables involving small numbers and the χ^2 test. J. Roy. Statist. Soc. Supplement 1 1, pp. 217-235.

1924

FISHER, R. A. (1924). The conditions under which chi-square measures the discrepancy between observation and hypothesis. J. Roy. Statist. Soc. 37, pp. 442-450.

1922

FISHER, R. A. (1922). On the interpretation of chi-square from contingency tables, and the calculation of P. J. Roy. Statist. Soc. 85, pp. 87-94.

1900

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos. Mag., Series 5 50, pp. 157-172.